

# Vowel Change in New Zealand English – Patterns and Implications

A thesis submitted in fulfilment of the  
requirements for the Degree of  
Doctor of Philosophy in Linguistics  
in the University of Canterbury  
by Christian Langstrof  
University of Canterbury  
2006

**To Katie Drager, a true hero!**

**Pierdo el tiempo pensando  
en lo esencial  
que a veces dejo pasar**

***Héroes del Silencio***

## **Acknowledgements**

None of the work this thesis is based upon would have been possible without the generous financial support from the Deutscher Akademischer Austauschdienst (DAAD) and the University of Canterbury, who both granted me scholarship covering the fees as well as living expenses during the entirety of my project.

As far as academic support is concerned, I am indebted to Jen Hay for her relentless support and her continued encouragement as my primary supervisor throughout the project. In addition, I would like to thank all the people at the Department of Linguistics at the University of Canterbury for providing a helpful and stimulating atmosphere during my stay here.

I would especially like to thank Lyle Campbell for his wholehearted support during the application procedure, which greatly influenced my decision to come to UoC in the first place.

The intermediate archive data was collected by Rosemary Goodyear, Lesley Evans, and members of the ONZE team. The work done by members of the ONZE project in preparing the data, making transcripts, and obtaining background information is also acknowledged.

Finally, I would like to thank my parents for their support in all things personal as well as financial over many years!

Christian Langstrof

## Table of contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Methodology</b>	<b>19</b>
2.1	Introduction	19
2.2	The speakers	19
2.3	The variables	21
2.4	Acoustic analysis	22
2.5	Normalisation	24
2.5.1	Why normalise?	24
2.5.2	A selection of formerly proposed normalisation procedures	29
2.5.3	Normalising intermediate speakers	31
2.5.4	Discussion	50
2.5.6	Conclusion	62
<b>3</b>	<b>The Short Front Vowels – KIT/DRESS/TRAP</b>	<b>63</b>
3.1	Introduction	63
3.2	Background	63
3.2.1	Etymological considerations	63
3.2.2	The SFVs in related varieties	65
3.2.4	The SFV shift in New Zealand English	66
3.3	Results – Overall patterns	71
3.3.1	Evidence for a push-chain scenario	78
3.3.2	The ‘split system’ in the early male sample	80
3.3.3	Discussion – Means and allophones	81
3.3.4	Factor analysis	85
3.3.5	Correlations between vowels for individual speakers	97
3.4	Lexical Frequency	102
3.4.1	Factor analysis	102
3.5	Discussion – SFVs in F1/F2 space	106
3.6	Vowel Duration	107
3.6.1	Background – Vowel duration in English	108
3.6.2	Vowel duration in the intermediate period	110
3.6.2.1	Vowels in Monosyllables	112
3.6.2.2	Vowels in Polysyllables	113
3.6.3	Factor analysis	115
3.6.4	Discussion – Vowel duration	120
3.7	Conclusion – The NZE Front Vowel Shift	126
<b>4</b>	<b>The Front Centring Diphthongs – NEAR/SQUARE</b>	<b>128</b>

4.1	Introduction	128
4.2	Background – Historical origins	128
4.3	Mechanisms of phonemic mergers	131
4.4	NEAR/SQUARE in the intermediate period – Static properties	133
4.4.1	General patterns	133
4.4.2	Factor analysis	142
4.5	Discussion – Nucleus position	147
4.6	Dynamic properties	148
4.7	Stress	152
4.8	Merger by approximation, transfer, or expansion?	154
4.9	Lexical frequency	162
4.10	Durational properties	165
4.11	Conclusion	169
<b>5</b>	<b>What is a FCD nucleus? – <i>On the (non)-correspondence between the nuclei of the front centring diphthongs and their monophthongal counterparts</i></b>	<b>171</b>
5.1	Introduction	171
5.2	Quality – Formant frequency averages and shapes of distributions	174
5.3	Duration revisited	183
5.4	Conclusion	186
<b>6</b>	<b>Broad A in intermediate NZE</b>	<b>188</b>
6.1	Introduction	188
6.2	Background	188
6.3	The START vowel in early NZE	192
6.4	Formant frequency distribution of the START vowel in the intermediate period	193
6.5	Merger by transfer?	199
6.6	Vowel duration	206
6.7	Lexical diffusion and the role of proper nouns	208
6.8	Conclusion	212
<b>7</b>	<b>The short front vowels – An exception to directionality in chain shifts?</b>	<b>214</b>
7.1	Introduction	214
7.2	Vowel types and subsystems	215
7.3	Change within and across subsystems	228
7.4	Front vowels on tracks	240
7.5	Conclusion	244

<b>8</b>	<b>Chain shifts in multidimensional vowel spaces – An evolutionary approach</b>	<b>246</b>
8.1	Introduction	246
8.2	A mechanistic model of vowel shifts	246
8.3	Vowel shifts in multi-dimensional vowel spaces	257
8.4	Are vowel spaces segmented?	262
8.5	Conclusions, caveats and outlooks	273
<b>9</b>	<b>A phonological afterthought - <i>Vowel change as optimisation: Why the NZE front vowel shift is not a good example</i></b>	<b>277</b>
9.1	Introduction	277
9.2	Optimisation in phonological systems	277
9.2.1	What is optimal?	277
9.2.2	Hawkins' analysis of NZE vowels	279
9.2.3	General problems with the concepts of 'optimality' in vowel systems	280
9.2.4	Problems with the concept of 'optimality' specific to NZE short front vowels	281
9.3	Discussion/Conclusion	286
<b>10</b>	<b>Conclusion</b>	<b>288</b>
	<b><i>References</i></b>	<b>295</b>

## Figures

- 2.1 Average raw frequency values of F1 against F2 of two speakers from the Intermediate Archive
- 2.2 Token scatter of DRESS and TRAP in the speech of 9 late Intermediate speakers
- 2.3 CART representations of pooled F1 data
- 2.4 CART representations of normalised formant frequency data in DRESS
- 2.5 CART representations of pooled F2 data
- 3.1 A schematic representation of the New Zealand English SFV shift
- 3.2 A schematic representation of the NZE front vowel shift
- 3.3 Two potential scenarios of a push-chain relationship between two vowels
- 3.4 F1/F2 averages of KIT, DRESS and TRAP
- 3.5 Mean frequency values of the first two formants of the lexical sets of KIT, DRESS and TRAP as well as their contextual variants
- 3.6 F1/F2 overlap in the speech of an individual speaker from the early male sample
- 3.7 Two possible scenarios for the NZE front vowel shift in the top two heights
- 3.8 (a) CART analysis of pooled F1 measurements of all three SFVs
- 3.8 (b) CART analysis of all KIT F1 measurements
- 3.8 (c) CART analysis of all DRESS F1 measurements
- 3.8 (d) CART analysis of all TRAP F1 measurements
- 3.9 (a) CART analysis of pooled F2 measurements
- 3.9 (b) CART analysis of all KIT F2 measurements
- 3.9 (c) CART analysis of all DRESS F2 measurements
- 3.9 (d) CART analysis of all TRAP F2 measurements
- 3.10 Individual mean values in those dimensions that are assumed to be interrelated
- 3.11 Individual mean values of DRESS F1 and KIT F2 before alveolar consonants

- 3.12 Individual mean values of formant frequency and year of birth
- 3.13 CART analysis of KIT F2 measurements including the independent variables *lemma frequency* (logged) and *word frequency*
- 3.14 Average durations of KIT, DRESS and TRAP before voiced and voiceless stops
- 3.15 Classification and Regression Tree of pooled duration measurements along the independent variables F1/F2/age/vowel/voice/syllable
- 3.16 Classification and Regression Tree analysis of duration measurements broken down into the three SFVs
- 3.17 Differential degrees of overlap between the lexical sets of KIT and DRESS in the speech of an early male speaker
- 3.18 F1/F2/duration plots of KIT, DRESS and TRAP in the early male and late female sample
- 4.1 Front vowel formant frequency averages in the Intermediate sample
- 4.2 Front vowel formant frequency averages in the Intermediate sample, broken down into age and speaker sex
- 4.3 Mean formant frequency and SD in six groups of Intermediate speakers
- 4.4 CART representations of major predictors of pooled NEAR/SQUARE data in the Intermediate sample
- 4.5 CART representations of formant frequency data from the early male and the late female group
- 4.6 Lowess line scatterplot smoother models of NEAR and SQUARE trajectories in the speech of 6 groups of speakers from the Intermediate Archive
- 4.7 Hypothetical token distributions of two vowels undergoing merger
- 4.8 Token distributions of pooled F1 data from the Intermediate sample
- 4.9 Token distributions of NEAR/SQUARE in the early male and the late female sample
- 4.10 Formant frequency plotted against lexical frequency in the Intermediate sample
- 4.11 Formant frequency plotted against lexical frequency for NEAR in the Intermediate sample



- 4.12 Classification and regression tree analysis of duration in centring diphthongs along the variables *F1/F2/age/syllable/voice*
- 4.13 NEAR/SQUARE distributions of 4 late female speakers
- 5.1 Formant frequency averages of FLEECE – NEAR – SQUARE – DRESS – KIT
- 5.2 Individual mean F1 values of DRESS and SQUARE
- 5.3 Two potential scenarios of DRESS and SQUARE raising under increasing correlation over time
- 5.4 Formant frequency values of NEAR and FLEECE in the early and late sample
- 5.5 F1/F2 plot of all tokens of NEAR, FLEECE and KIT in the early male sample
- 5.6 Average time-normalised formant frequency trajectories of NEAR and SQUARE in the Intermediate sample
- 5.7 Proportion-through-trajectory measure, schematised
- 6.1 Formant frequency plots of DANCE/ BATH against TRAP/START averages
- 6.2 Classification and regression tree analysis of pooled F2 data on the independent variables *Manner of articulation of the following sound/Speaker Sex/Age*
- 6.3 Classification and regression tree analysis of pooled F1 data on the independent variables *Manner of articulation of the following sound/Speaker Sex/Age*
- 6.4 Formant frequency plots of DANCE/BATH against TRAP/START averages of the early male and late female sample
- 6.5 Histogram representations potential of broad A words
- 6.6 Histogram representations of BATH
- 6.7 Histogram representations of DANCE
- 6.8 Formant frequency position of the lexical item ‘dance’ vs. all other pre-nasal tokens
- 6.9 Formant frequency plots showing conservative realisations of place names in the speech of an early female, a medium female and a late male speaker
- 7.1 Hypothetical vowel space comprising 2 subsets and 3 heights
- 7.2 A two-track model of vowel space

- 7.3 A diagram of a hypothetical tense/lax vowel space
- 8.1 Formant frequency plots showing tokens of DRESS both in monosyllables before voiced stops as well as their complementary environments

## Tables

2.1	Speakers analysed in the present thesis and their birthdates
2.2	Formant frequency averages and standard deviations of DRESS and TRAP in the late sample
2.3	Hotelling-Lawley trace scores between overall distributions of DRESS and TRAP as well as male and female distributions within one phoneme
2.4	Average Euclidean distances between all tokens of DRESS and TRAP to their own mean and the mean of the respective other vowel distribution
2.5	Ratios of average Euclidean distances of DRESS and TRAP tokens
2.6	F2 in hz and token numbers of KIT, TRAP and STRUT in all six subsamples
2.7	Spearman correlation coefficients and significance levels in raw data and normalised data
2.8	Wilcox test results and significance levels for all speakers from the Intermediate sample
2.9	Correlation test results between log-transformed F1/F2 and significance levels for all speakers from the Intermediate sample
3.1	Euclidean distances between the mean values of the lexical sets of KIT, DRESS and TRAP in the speech of the six groups of Intermediate speakers
3.2	Coding categories
3.3	Correlation coefficients and their significance levels
3.4	Correlation coefficients and significance levels between word frequency/lemma frequency and F1/F2 for all three SFVs
3.5	Correlation coefficients and significance levels between word frequency/lemma frequency and F1/F2 for all three SFVs; excluding function words
3.6	Token numbers of the words <i>if</i> and <i>in</i> , broken up into subsamples
3.7	Duration ratios between different vowels before voiced and voiceless stops in American English
3.8	Number of tokens obtained from two groups of Intermediate speakers
3.9	Token numbers obtained from two groups of Intermediate speakers

broken down into occurrences in monosyllabic and polysyllabic words

- 3.10 Average durations in milliseconds (ms) of short front vowels in the speech of 9 intermediate speakers
- 3.11 Correlation coefficients between the variables *duration* and *F1/F2* in the SFVs
- 3.12 Hotelling-Lawley trace scores of the distributions of KIT, DRESS and TRAP in the early male and the late female sample
- 4.1 Hotelling-Lawley traces and corresponding significance levels for NEAR - SQUARE distributions in six groups of Intermediate speakers
- 4.2 Average formant frequencies and standard deviations of the nuclei of NEAR and SQUARE in each of the six groups of Intermediate speakers
- 4.3 Mean formant frequency values, standard deviations and 1<sup>st</sup> and 3<sup>rd</sup> quartiles in the offglide of NEAR and SQUARE in the speech of six groups of Intermediate speakers
- 4.4 Hotelling – Lawley scores and significance levels for /ɪə/ - /εə/ distributions in the speech of six groups of Intermediate speakers, resolved for three levels of stress
- 4.5 Average formant frequencies and standard deviations of the nuclei of NEAR and SQUARE in each of the six groups of Intermediate speakers
- 4.6 Correlation coefficients and significance levels between formant frequency and lexical frequency
- 4.7 Duration averages of the front centring diphthongs in two Intermediate speaker groups
- 5.1 Hotelling-Lawley trace scores and significance levels between the distributions of FLEECE – NEAR – SQUARE – DRESS – KIT
- 5.2 Correlation coefficients and significance levels between the individual speaker averages of FLEECE – NEAR – SQUARE – DRESS – KIT
- 5.3 Hotelling-Lawley trace scores for distributions of DRESS and SQUARE
- 5.4 Hotelling-Lawley trace scores for DRESS and SQUARE for each individual Intermediate speaker
- 5.5 Hotelling-Lawley trace scores for distributions of NEAR and FLEECE
- 5.6 Hotelling-Lawley trace scores and significance levels between FLEECE, NEAR and KIT in the early male sample

- 6.1 Present-day reflexes of ME /a/ in different varieties of English
- 6.2 Duration differences and significance levels of pooled data from the Intermediate sample
- 7.1 Transcription of Standard English vowels in IPA, Wells (1982) and Trager/Smith (1957)
- 8.2 Favoured variants of KIT and DRESS tokens in the evolution of vowel systems of the type shown by the early Intermediate speakers
- 8.3 Hypothetical front vowels and F1/F2/duration values
- 8.4 Euclidean distances between six hypothetical vowels
- 8.5 Euclidean distances between the six hypothetical vowels as well as two upgliding diphthongs along the dimensions F1/F2/duration/diphthongisation
- 8.6 Euclidean distances between KIT, DRESS and TRAP in the early male sample along the dimensions F1/F2/duration

## Abbreviations

AusE	Australian English
EModE	Early Modern English
FCD	Front Centring Diphthong
GenAm	General American (i.e. American English in the sense of Wells (1982))
GVS	Great Vowel Shift
ME	Middle English
MHG	Modern High German
MHGLVS	Middle High German Long Vowel Shift
NZE	New Zealand English
OE	Old English
RP	Received Pronunciation
SAE	South African English
SFV	Short Front Vowel(s)
SPE	The sound pattern of English (i.e. Chomsky and Halle (1968))

# Thesis title: Vowel Change in New Zealand English – Patterns and Implications

## Abstract

This thesis investigates change in a number of phonological variables in New Zealand English (NZE) during a formative period of its development. The variables under analysis are the short front vowels /ɪ/, /ɛ/, /æ/, the front centring diphthongs /ɪə/ and /ɛə/, and the so-called ‘broad A’ vowel. The sample includes 30 NZE speakers born between the 1890s and the 1930s (the ‘Intermediate period’).

Acoustic analysis reveals that the short front vowel system develops into one with two front vowels and one central vowel over the intermediate period via a push chain shift. There is evidence for complex allophonisation in the speech of early intermediate speakers. I argue that duration plays an important role in resolving overlap between vowel distributions during this time.

With regard to the front centring diphthongs there is approximation of the nuclei of the two vowels in F1/F2 space over the intermediate period as well as incipient merger in the speech of late intermediate speakers. Although the merger is mainly one of gradual approximation, it is argued that patterns of expansion of the vowel space available to both vowels are also found.

The analysis carried out on the ‘broad A’ vowel reveals that whereas flat A was still present in the speech of the earlier speakers from the sample, broad A had become categorical toward the end of the intermediate period. It is shown that, by and large, the process involves discrete transfer of words across etymological categories.

The final chapters discuss a number of theoretical implications. Processes such as the NZE front vowel shift suggest that a number of previously recognised concepts, such as ‘tracks’ and ‘subsystems’, may either have to be relaxed or abandoned altogether. It is argued that chain shifts of this type come about by rather simple mechanisms that have a strong resemblance to functional principles found in the evolution of organisms. A case for ‘fitness’ of variants of a given vowel will be made. Phonological optimisation, on the other hand, is not a driving force in this type of sound change.

1.

## Introduction

This thesis is about sociophonetics. Its major goal is to document change in a number of phonological variables in New Zealand English by presenting results of detailed acoustic analysis. In addition, I will discuss both theoretical implications of the data as well as methodological questions.

The variables that were chosen for the analysis are the short front vowels /ɪ/, /ɛ/ and /æ/, the front centring diphthongs /ɪə/ and /ɛə/ and the so-called ‘broad A’ vowel, i.e. the distribution of words over two different etymological vowels (/æ/ and /a/). It has previously been shown that these variables set off contemporary New Zealand English from its closest cousins (Australian English, South African English, and Southern British English).

The sample which the data was drawn from consists of 30 speakers from the Intermediate Archive, held at the ONZE (Origins of New Zealand English) lab at the University of Canterbury. These speakers were born between the 1890s and the 1930s and were selected for two reasons: Firstly, no detailed acoustic analysis had previously been carried out on that corpus, whereas this was done for both earlier and later speakers. Therefore, the research presented here fills a void in the documentation of New Zealand English. Secondly, previous research indicated that the Intermediate period may have been a crucial stage in the development of contemporary New Zealand English, at least with respect to the aforementioned variables. By comparing their data from first and second generation speakers of New Zealand English to the current pronunciation, Gordon et al. (2004) found a great degree of discontinuity. That is, none of the processes that transformed the vowel system of NZE to its current state were completed in the speech of the first- and second generation of New Zealanders.

Given the availability of sound recordings of speakers of a given variety over its entire history, and given the empirical analyses that provide us with detailed information about the above variables over the entire history, we are in a position to draw some



conclusions of more general scope for theories of language change and dialect formation. Substantial parts of this thesis are dedicated to these issues.

Having said in the initial paragraph that this thesis is about sociophonetics, I would like to point out that this is only part of the story. The methodological approach to the empirical data is sociophonetic, and this constitutes the main part of this thesis (chapters 3-6). The approach taken in these chapters is to present results from formant frequency measurements of a large number of vowel tokens using phonetics software, and to process these results using regression modelling, t-tests, and the like. I pointed out above that the empirical findings presented in this thesis are to be regarded as a contribution to our knowledge about the history of New Zealand English. However, students of a history of a given variety rarely stop at some particular point in time, but would rather like to see any such research embedded in a larger picture in both historical as well as theoretical/explanatory terms, a goal I pursue in chapters 7-9.

Therefore, this thesis should be regarded as a ‘inter-subdisciplinary’ work that bears on sociophonetics, historical linguistics, epistemology and phonology.

### *The structure of this thesis*

## Chapter 2 – Methodology

Chapter 2 gives information about the speaker sample (age, grouping into subsamples) as well as a careful description of the formant frequency measurement procedures that were used.

In addition, I will present a lengthy discussion on speaker normalisation, a topic that has come up during the analysis of the raw measurements. Different data pools from the sample will be drawn upon to examine three commonly used normalisation procedures (Neary’s log-mean method, Lobanov’s z-score method, and Gerstman’s range method). The purpose of this discussion is twofold: Firstly I will try to justify the method used throughout the remainder of this thesis. In addition, it should serve as a illustration of the type of bias introduced by subjecting raw formant frequency data to

different kinds of normalisation procedures. It will be concluded that although normalisation yields useful results for sociolinguistic purposes, normalised data always need to be handled with caution, and the limitations of different normalisation methods have to be spelled out prior to presenting results based on this data.

The somewhat pessimistic conclusion of this discussion is that, for the purposes of studies similar to the one presented in this thesis, ‘perfect’ normalisation is both necessary as well as impossible. A more appropriate methodology to tackling large bodies of data from heterogeneous corpora might be present relative distances between vowel distributions across different speakers.

### Chapter 3 – The Short Front Vowels KIT, DRESS and TRAP

Chapter 3 presents an analysis of both formant frequency and duration measurements of the vowels /ɪ/, /ɛ/ and /æ/. Data from approximately 4500 tokens is presented. These vowels have undergone a quality change in New Zealand English, where /ɛ/ and /æ/ have risen, whereas /ɪ/ centralised, but the exact interrelations between them have long been a matter of debate. It is shown in this chapter that it is during the intermediate period that the contemporary short front vowel system develops in NZE. Whereas the most conservative subsample (early male speakers) show a rather conservative set-up of three front vowels of different heights, later speakers show the modern system of two front vowels and one centralised vowel. Conclusive evidence for a push-chain relationship will be presented. It will also be shown that early speakers show complex allophonisation in /ɪ/, where both centralised and raised allophones are found. Non-parametric regression modelling reveals broad phonetic conditioning within /ɪ/ as well as allophonic interdependencies between vowel distributions. A case is made for extensive overlap between /ɪ/ and /ɛ/ in F1/F2 space in the speech of individual speakers, which shows that the chain shift comes about sequentially, i.e. one step does not imply the following one in the phonology of individuals.

A further step in the analysis discusses the role of vowel duration in the chain shift. It will be shown that there are substantial differences between the short front

vowels with respect to both absolute duration and the degree of allophonisation. It will be demonstrated that those vowels that show similar trajectories of movement over time in F1/F2 space (i.e. /ɪ/ and /ɛ/) are also more similar to each other in the duration domain than either one of them is to the centralising vowel /ɪ/. In addition, it will be shown that the most extensive degree of durational allophonisation is found in /ɛ/.

The conclusion will be that duration may have been an important dimension in the chain shift as it can be shown to disambiguate overlapping distributions. Therefore, the concept of ‘short front vowels’ is argued to be inadequate in the case of /ɪ/, /ɛ/ and /æ/ in NZE.

#### Chapter 4 – The front centring diphthongs /ɪə/ and /ɛə/

A phonemic merger between the front centring diphthongs (FCDs) /ɪə/ and /ɛə/ is the topic of chapter 4. Many current speakers of NZE have a merger of these two elements (usually on the higher vowel), whereas they remain distinct for speakers of related varieties (as well as a number of NZE speakers). It has previously been shown that the two elements were clearly distinct in the speech of first- and second-generation NZers, the merger must therefore have occurred afterwards.

It will be shown that the intermediate period exemplifies what one might call a ‘pre-merger’ state with increasing similarity between the two phonemes. As the distributions of the two vowels approach each other, there holds an intermediate stage where an artefactual unimodal distribution arises via the adding up of outliers, which might have been the crucial change-over point in the cognitive representation of the two vowel phonemes. It is also shown that for late speakers, the distance between the two distributions is smaller (and the difference less significant) in contexts of high word stress, which is regarded as further evidence of the vowels developing towards a merged state over the intermediate period. A more tentative case will be made for extension of the vowel space available to both phonemes, which would make the merger one of both gradual approximation as a result of both raising of the nucleus of /ɛə/ as well as expansion of the vowel space occupied by the nucleus of /ɪə/. In addition, it will be

shown that at the initial stages of raising of the nucleus of /εθ/, there seems to have been a transitional reactionary movement of the nucleus of /ɪθ/ towards closer/fronter articulation, i.e. an ephemeral chain-shift.

Finally, duration measurements are also discussed and it will be shown that the duration differential that we observed between /ɪ/ and /ε/ in chapter 3 does not hold in the case of the nuclei of /ɪθ/ and /εθ/. This has implications for the conceptualisation of front vowel systems, which are discussed in chapter 5.

## Chapter 5 – What is a FCD nucleus?

The results presented in chapter 4 indicate that the relationships between the front monophthongs and the front centring diphthongs are complicated and require a discussion, which is the topic of chapter 5. The most important questions in this context are whether the nuclei of the front centring diphthongs are directly equivalent to elements from the monophthong system and if so (as is generally assumed in both phonology and historical linguistics on the basis of historical origin and/or synchronic alternation patterns), what the correspondences are.

It is argued that for these particular speakers, the patterns are far from clear. As for the relationship between /ε/ and the nucleus of /εθ/, it will be shown that the two elements correlate more strongly in the speech of later speakers. What is less clear is whether this is due to a ‘real’ association of the two in the later sub-samples, or whether this is an artefactual effect of different processes occurring at the same time in the speech of these individuals (i.e. /ε/ raising and /εθ/ raising).

The case of the nucleus of /ɪθ/ is more complicated. Common textbook and dictionary transcriptions indicate identity with /ɪ/ (i.e. transcribed as /ɪθ/). Although this might be adequate for other varieties of English, it is hardly tenable for contemporary NZE (where /ɪ/ is a central vowel while the nucleus of /ɪθ/ remains front). The speakers analysed in this thesis do not show congruence between /ɪ/ and /ɪθ/ at any stage, with the difference increasing over time. Additional evidence against equating the nucleus of /ɪθ/ with /ɪ/ is provided by data from duration measurements.

Although a case can be made for equating the nucleus of /ɪə/ with /i/, the evidence is not perfect either.

## Chapter 6 – Broad A in Intermediate NZE

Chapter 6 looks at the distribution of words across two different etymological vowels in the Intermediate period. Lexical items such as ‘dance’ and ‘half’ show different vowels in different varieties of English in that some accents (especially in North America) show a front vowel [æ], whereas others (RP/contemporary NZE) show a central or back vowel [a] ~ [ɑ]. In addition, certain varieties show variable realisations of this vowel (e.g. AusE).

It will be shown in chapter 6 that the change from /æ/ to /a/ was not yet completed at the early intermediate stage. Rather, speakers from this period retain ‘flat A’ (i.e. the front variants) in some words, especially in front of nasal consonants. The most innovative intermediate speakers show broad A across the board. Distributional data indicates that, by and large, the change involves transfer of lexical items across the two vowels. However, the patterns are less clear than a pure transfer scenario would lead one to expect, as we find non-matching distributional maxima in potential broad A words compared to non-broad A instances of /a/.

Evidence from duration measurements suggests a similar pattern: Whereas on the whole, vowels in broad A words that retain [æ] show duration values similar to other instances of /æ/, there are differences in detail. In pre-fricative environments, words that contain [a] show a slightly (and statistically significant) duration difference to pre-nasal tokens, a pattern opposite to what one would expect on phonetic grounds. Thus, although the overarching mechanism is one of transfer of word class membership from /æ/ to /a/, the new /a/ distribution retains fine-grained phonetic differences that reflect etymological origin.

The special role of proper nouns in this process will also be discussed, and it will be shown that for some speakers at least, proper nouns are more susceptible than common nouns to retain the conservative variant.

## Chapter 7 – The short front vowels – An exception to directionality in chain shifts?

Chapter 7 discusses the status of the short front vowels within the framework of Labov's (1994) theory of directional vowel change. It has been pointed out that these elements violate a number of principles of that theory, in that they behave in a non-uniform fashion across different varieties of English. For example, both /ɛ/ and /æ/ underwent raising and fronting in the southern Hemisphere varieties, whereas /ɛ/ lowers while /æ/ raises in the Northern Cities shift (cf. the extensive documentation in Labov (1994/2001), and M. Gordon (2001)). Recent evidence suggests lowering of /æ/ in contemporary AusE. These processes are unexpected within Labov's theory, which essentially states that vowels of a given subsystem behave in a uniform fashion, where long (or 'tense') vowels rise while short ('lax') vowels fall.

The discussion in chapter 7 will attempt to reconcile the observed processes affecting the short front vowels of English with the overall regularities found in other languages as well as English non-front vowels. It is argued that a number of processes affecting the front vowel system in EModE have brought about a highly untypical front vowel space in most standard varieties of Modern English. More specifically, the merger of ME /e:/ and /ɛ:/ on a high vowel /i:/ have left the front vowel space of English bereft of a long monophthong. In addition, the lowering of ME /u/ and the subsequent fronting of ME /a/ to [æ] introduced a further asymmetry by adding another short vowel to the front series. As a result, the front vowel system of English lacks clear reference points for assigning length, peripherality, and tenseness. Therefore, subsequent histories of English are free to take different paths with regard to these elements. The major implication for the theory of directionality in vowel change is that one cannot rely on etymological status nor on phonotactic criteria alone in assigning categories like 'tenseness' or 'peripherality' to members of a vowel system. Rather, such labels need to be reassigned depending on the historical contingencies that affect different varieties of English.

This view makes a number of empirical predictions with regard to the short front vowel data in NZE. If a given segment is hypothesised to have undergone

reinterpretation as a member of another subsystem, one would expect this element to exhibit properties that are more similar to other members of that subsystem compared to an etymologically comparable vowel that has not undergone this reinterpretation. This is what we find in the intermediate data: Compared to /ɪ/, which undergoes centralisation and lowering in NZE, both /ɛ/ and /æ/ show significantly higher duration values. In addition, both show a higher degree of durational allophonisation, a property typically associated with long vowels (cf. the discussion/references in chapter 3). Therefore, a case is made for /ɛ/ and /æ/ having undergone reinterpretation as long vowels in NZE, while /ɪ/ remains a short vowel. If this view is accepted, their behaviour is indeed in line with Labov's theory.

## Chapter 8 – Change in multidimensional vowel spaces – An evolutionary approach

Chapter 8 goes one step further and claims that some of the primitives advocated in Labov's theory are redundant. More specifically, it is argued that notions such as 'track' and 'subsystem' should be abandoned. The main tenet is that chain shifts come about by functional mechanisms similar to those found in biological evolution, namely variation and selection. It is argued that language learners select those variants of vowel distributions that signal the identity of that vowel best, i.e. are least susceptible to be confused with an instance of another vowel. Vowel distributions move toward regions in vowel space that were formerly occupied by favourable variants. Therefore, vowel distributions seek to achieve local distance equilibria.

An important point in this context is that distance is a function of more than two phonetic dimensions. Rather, other dimensions such as duration and diphthongisation contribute to the location of a given vowel in vowel space. What matters in chain shifts are distance relations between vowels along all relevant dimensions. This view circumvents problems encountered in accounting for chain shifts under more conventional views without having to posit ad-hoc concepts such as 'subsystems' or 'tracks'.

The view as advocated in chapter 8 recognises vowel spaces in their entirety without any subdivisions. That is, any vowel system is a discontinuous distribution of phonetic properties on a continuous vowel space landscape. Any region of the overall vowel space is in principle ‘inhabitable’ by any vowel distribution as a result of vowel movements. Formerly proposed concepts such as ‘tracks’ and ‘subsystems’ are claimed to be epiphenomena of these more basic functional interdependencies between vowel distributions. It will be claimed that these simple mechanisms are sufficient to predict the directionalities in chain shifts as observed by Labov while at the same time being able to incorporate ‘aberrant’ processes such as the Southern Hemisphere front vowel shifts (as well as the contemporary interaction between /ɛ/ and /i:/).

A tentative calculation of distance ratios between the short front vowels based on the data from the intermediate speakers is presented, and it will be shown that the predictions are largely borne out. It is shown that locations in vowel space that are occupied by favourable variants form the target locus of the subsequent development. In addition, it is shown that allophones of comparatively low fitness undergo change at a faster rate.

## Chapter 9 – Vowel change as optimisation – Why the NZE front vowel shift is not a good example

A specific proposal regarding phonological implications of the short front vowel shift in NZE is taken up in chapter 9. Hawkins (1976) argues that the shift is an example of phonological optimisation in vowel systems. On his view, the input system of short vowels consists of three front vowels, one central vowel and two back vowels. This set-up is asymmetrical as well as uneconomic in terms of feature matrices. The modern system, however, is optimised since it consists of only two contrastive heights as well as taking advantage of any possible constellation of features. That is, those feature combinations that are necessary in order to fully describe the short vowel system of current NZE are at the same time sufficient.



Chapter 9 argues that this idea is untenable on both empirical as well as theoretical grounds. Empirically, data analysed in chapter 3 will be drawn upon to show that the chain shift came about sequentially and gradually. That is, early intermediate speakers show a setup of front vowels with /ɛ/ being rather close to /ɪ/, which implies that raising of the former does not necessarily imply full centralisation of the latter. In order to satisfactorily define such systems in terms of SPE features, one would have to resort to rather elaborate allophonic machinery. Therefore, one would have to posit a generation of speakers that sacrifices a relatively uncomplicated input system for an awkward ephemeral one in order to allow for an optimal system to arise in the speech of their successors, which seems unlikely on theoretical grounds. In addition, comparative data from related varieties of English show that the ‘optimal’ system isn’t particularly widespread.

### *Some background assumptions*

Of course, no research project can possibly fulfil the positivist ideal that the data as such is neutral and theories emerge solely as results of objective theory-free assessments of a data set. It therefore seems prudent for anyone to lay their cards on the table in advance of putting forward a large-scale research project such as the present one. Apart from a general commitment to ‘standard positivist’ views (such as *modus ponens* is a valid mode of implication, or Occam’s Razor is a reasonable way of evaluating conflicting views of the same ‘correctness’, that viable theories are falsifiable rather than verifiable (cf. Popper 1982, Lass 1980), I presuppose a number of rather more specific points which emerge from the study of phonetics, psychology and linguistics over the last ~ 150 years and which have a bearing on the research in this paper in both its methodology and its conclusions. Each of these presuppositions deserves a brief comment.

#### *1. Formant frequency relates to (among other things) frontness and height of vowels.*

And, by implication:

*I'. Acoustic analysis of energy concentrations in vowel spectra is a valid way of scientifically assessing vowel quality.*

These two statements are obviously intertwined and relate to the justification of the basic method of data analysis in this paper, namely acoustic analysis of vowel formants. At least since the early research by Liberman and others at Haskins Laboratories (summed up in Liberman (1996)) it has been clear that perceived vowel quality is primarily a function of the location of concentrations of acoustic energy in the overall spectrum of a vocalic utterance. What has been less clear is how *linear* the correspondence between frequency of emission and perceived frequency is and how integrative the perception mechanism is. The linearity problem relates to the fact that in speech perception, equal increments in frequency of emission are not perceived as equal above a certain frequency range (the relation is quasi-logarithmic, cf. Syrdal and Gopal (1986), Moore and Glasberg (1983), Zwicker and Terhardt (1980)). This does not, however, have any bearing on the basic representation of two given vowel articulations *relative to each other*, which will be the main 'mode' of reasoning in this paper. That is, if vowel token A has a higher F1 in hz than vowel token B, it does so irrespective of whether the values are plotted onto a hertz scale or a BARK/ERB scale. It is for this reason that this paper states values in hertz throughout.

The question of perceptual integration relates to which spectral properties contribute to vowel quality in the first place. Although it seems clear that the first two formants do contribute, the role of higher formants as well as that of the fundamental frequency of phonation is less clear. As for the former, various researchers (cf. Chistovich (1979), Johnson (1989) and the application in de Boer (2001)) assume that higher formants (especially F3) do play a role under certain circumstances, especially if the location of the second formant is relatively high (as is the case for front vowels). More specifically, if a critical minimum threshold of separation between F2 and F3 is not met, the perceptual correlate is a weighted average of the two (the critical threshold seems to be around 3.5 BARK, cf. Chistovich 1979).

In addition to this, the role of the fundamental frequency of phonation is not clear. Whereas traditional source-filter theory (Fant 1960) assumes independence between the source ( $f_0$ , i.e. the rate of vibration of the vocal folds) and the filter (formant frequencies as a function of ‘filter size/shape’), a number of researchers (Miller (1953), Fujisaki and Kawashima (1968), Johnson (1990)) assume some interdependence between the two.

However, since these matters do not seem to be sufficiently settled, I will adhere to the conservative view that the location of the first two formants are the major correlate of vowel frequency<sup>1</sup>.

I will therefore equate frequency values of the first two formants with vowel quality throughout the paper, whereby F1 correlates inversely with height, and F2 correlates with frontness.

## *2. Apparent time mirrors real time*

This assumption is important since it relates to both the methodology as well as the conclusions drawn from the analysis. In addition, it is to some extent contestable and therefore requires some discussion. The terms ‘apparent time’ and ‘real time’ are sociolinguistic constructs (cf. the discussion in Labov 1994: chapter 2) and relate to the mapping of synchronic variation to diachronic processes. More specifically, this tenet can be circumscribed as follows

*2'. If, at some specified point in time, two speakers of different age show different realisations of some linguistic variable X, this difference reflects change in X in that X changes from a realisation shown by the older speaker toward that of the younger speaker. **Ceteris Paribus**<sup>2</sup>.*

---

<sup>1</sup> Most of the dispute regarding perceptual linearity and integration seem to relate to the mapping of articulation to perception and may therefore be regarded as irrelevant to the present thesis, which is after all concerned with production only. However, if it can be conclusively shown that two vowel utterances which differ in the F2 dimension if plotted onto an F1/F2 chart would not do so under an integrative view, the two views would be (relevantly) at odds with each other since a statement such as ‘Vowel token A is fronter than vowel token B’ can hardly be assumed to have any import to a general analysis of vowel quality if that difference cannot possibly be perceived by anyone. After all, even if small-scale variation in the realisation of vowels is meaningful to the description of vowel systems and change, this variation must be acoustically mediated between speakers and therefore is required to be acoustically salient in the first place.

<sup>2</sup> *Ceteris paribus* refers here to an idealised constellation where other factors that are known to influence a speakers realisation of a given linguistic variable, such as speaker sex and social class, are perfectly controlled.

This works because of a further assumption:

*2''. Adult phonologies are relatively stable.*

With respect to the present thesis, this means that since all the interviews were carried out at approximately the same time, and since the interviewees are of different age, variation (at least directional variation) can be interpreted as reflecting successive stages of development in the variables under analysis, given that *ceteris paribus* is met. Although it has been shown that assumption 2'' is not strictly true and that adult phonologies do in fact change (Harrington et al. (2000 (a), (b))<sup>3</sup>), too much has been made out of this with regard to the validity of the 'apparent time' approach. More specifically, what has been shown is that if adult phonologies change, they do so in the direction of a later norm. This, however, is a fundamental problem only if two speakers of different age have the *same* realisation of a given variable. In this scenario (and under the assumption of adult speakers adopting later norms), we would not be able to infer that that variable has been stable over time. On the other hand, if two speakers of different age show *different* realisations of a variable, the worst that can happen is that we underestimate the *degree* of change, but not the *fact* of change itself nor the *direction* of change.

I will therefore assume that systematic differences between age groups reflect language change.

*3. There are two types of vowel change; a) gradual (neo-grammarian) and b) discrete (lexical transfer/diffusion).*

*3'. They are fundamentally different in both their mechanism as well as their 'phonological scope'.*

*3''. (a) is more common than (b).*

---

<sup>3</sup> Although it is not clear whether the changes documented by long-term studies on the pronunciation of individual speakers reflect a change in their phonology, or whether these individuals shift between different registers.

Assumption (3) essentially states that the most common type of change in vowels is that of slow gradual change over time. This type of change operates by and large on the level of the phoneme whereby all words containing a given vowel retain their etymological identity at all stages before, throughout and after some type of change (such as a vowel shift). This is the mechanism of vowel change (or more generally, ‘sound change’) as envisaged by the neogrammarian school of linguistics in the late nineteenth century (for an exposition of the basic principles cf. Brugmann (1888)), although in principle, this goes back to the early Indo-Europeanists, from Jones (1786) onwards). The understanding that in many cases etymological identity of a sound is preserved even after substantial sound change was the earliest triumph of systematic linguistics and allowed researchers to infer language histories and relations on the basis of synchronic data. The assumption that this is the default mode of change is still the working principle in comparative linguistics.

However, from the early 20<sup>th</sup> century onward, linguists working in dialectology have put into question the mechanism of phonemic sound change and claimed that the domain of sound change is the lexical item rather than the phoneme (or a principled subset thereof), summarised in the famous dictum ‘chaque mot a son histoire’. It was especially after the research by Wang and his collaborators (Wang (1977)) where the idea of the word as the fundamental unit of sound change became reasonably popular and was christened the ‘lexical diffusion’ hypothesis. However, studies of change in progress as well as reconsideration of some of Wang’s data have shown that although lexical diffusion exists, gradual sound change of the neogrammarian type is considerably more common, especially in vowel change (cf. ch. 18 of Labov 1994).<sup>4</sup>

The differences between the two types of processes are discussed in Labov (1994: 543) and can be summed up as follows:

- Vowel change by lexical diffusion is discrete, i.e. if a vowel changes from A to B, words containing etymological A will have either A or B in an in-between stage, but no intermediate realisations.

---

<sup>4</sup> What complicates this bifurcated picture of neo-grammarian incremental sound change vs. the dialectologists’ lexical word-by-word diffusion is the fact that lexical transfer was indeed recognised by the neogrammarians, but attributed to other principles such as systemic analogy (cf. the discussion in Kiparsky (1995)). It is not clear to what extent analogy plays a role in more recent lexical diffusion approaches.

- Vowel change via lexical diffusion is ‘anarchic’ in that there are no exactly specifiable phonological conditioning factors.
- Vowel change via lexical diffusion frequently splits etymological categories and is therefore ‘incomplete’.
- Vowel change via lexical diffusion operates on a ‘higher level of abstraction’, which is to say that it is only via non-gradual lexical transfer that a vowel can move to another subset of a complex vowel system, for example.
- For vowel change of the neogrammarian type, the inverse holds (i.e. It is gradual, works on the level of the phoneme, ‘exceptions’ are phonologically well defined (‘Verner’s Law’ (Verner (1875))), and generally involve only the spatial dimensions such as raising/lowering and backing/fronting/breaking, but not, for example, phonological lengthening).

Much debate has been raised over what is essentially an empirical question, and I will show in this paper that both types of change are exemplified by the data analysed here.

### *Conventions*

It has become a matter of convention in studies on NZE phonology to refer to vowels by using lexical set names, which goes back to Wells (1982). I will, by and large, continue this usage in the present paper. However, although this usage is not without its merits, it is both sufficiently idiosyncratic as well as problematic to deserve a brief discussion. The idea behind devising lexical set names for vowels is that it allows for making statements about etymological categories without running into a number of epistemological difficulties associated with phonemic IPA notation, notably the issue of how faithfully a given IPA notation (such as /i/) should reflect some articulatory reality. By using lexical set names (e.g. FLEECE), this problem is circumvented, as this approach is entirely neutral to the articulatory specifics of a given variety of English. Therefore, the reader is enabled to relate the segment to the appropriate value of their own dialect. While this approach allows for dialect-neutral statements regarding vowels as etymological/structural entities, it is problematic with regard

to structural change. More specifically, it is less straightforward if applied to items that are known to have undergone a change in class membership. This becomes clear in Wells' approach to 'broad A' words<sup>5</sup>: At some stage in EModE, a ME short central vowel fronted and developed long and short allophones [æ] and [æ:]. Furthermore, the long allophones underwent backing to [a:] in a number of varieties, which now have vowels of different quality and quantity in words like 'trap' and 'path'. The later stages (lengthening and backing) of this process are tantalisingly semi-regular in that although it seems to be broadly conditioned (vowels followed by /f θ ns nd mpl/, i.e. fricatives and nasal-initial consonant clusters), it is also subject to a large degree of lexical idiosyncrasy (cf. 'sample' with broad A vs. 'ample' with flat A). In addition, different varieties have stopped at different stages of the overall process. For instance, the first two stages (fronting and lengthening) are restricted to Southern British English and its extraterritorial derivatives. The second stage (backing) is restricted to Southern British English and its *late* extraterritorial derivatives, i.e. excluding North American varieties outside New England. Furthermore, AusE represents a further in-between stage in that backing occurs across the board before fricatives (i.e. /a:/ in words like 'path/pass/half'), but not necessarily before nasals (i.e. words like 'dance' can have either [æ:] or [a:] in AusE, subject to both sociolinguistic and regional conditioning, as well as individual idiosyncrasy).

We therefore have alternation of words between vowels in both diachronic and geographical terms. Wells' approach to this is to devise subset designations such as PATH and DANCE. The former term covers pre-fricative broad A words as well as their *etymological correlates* in flat A varieties; the latter covers pre-nasal environments. The problem here is as follows: Given that, in a synchronic perspective at least, a number of words show fluctuation between *pre-existing* categories (i.e. /æ/ and /a:/, 'TRAP' and 'START'), no variety in isolation can have a lexical set of PATH and DANCE. That is, these subset designations are post-hoc generalisations across the history of English as well as different outcomes of that history (i.e. different varieties). Therefore, no single speaker has a lexical set (or subset) of PATH or DANCE as a phonological entity that is somehow distinct from some other one.

An additional complication is the synchrony/diachrony interface: When does a given lexical set start to exist, and how do we relate this to its former stages? This is a practical

---

<sup>5</sup>Here I restrict myself to a brief outline of the historical issues involved, a more thorough discussion of this is found in chapter 6 of the present thesis.

problem which I faced in the current thesis, since I frequently refer to vowels in their historical context, often going back to a pre-modern stage (i.e. EModE, ME, OE). Consequentially, one would have to designate lexical set names for ME vowels as well, which is problematical in two respects: First of all, and somewhat trivially, no contemporary reader would be able to relate a ME lexical set name to some phonological entity of their own. In addition, and less trivially, it is unclear to what extent historical continuity can be encoded in such an approach. To give an example: The relation between the current lexical set of FLEECE and its ME predecessors is as follows: ME long /e:/ and /ɛ:/ merge on /i:/. Therefore, FLEECE comprises two ME phonemes. In addition, a number of allophones of ME /ɛ:/ are exempt from this process, mostly in pre-rhotic environments (cf. ‘pear’, ‘bear’, but also the famous ‘break’ and ‘great’, cf. Dobson (1957) and the discussion in Labov (1994, part C)). In addition, shortening of a number of ME /ɛ:/ words also affected the relation between ME /ɛ:/ and its modern reflex, in that words like ‘head’ and ‘dead’ are not members of FLEECE.

Bearing these considerations in mind, how can we designate ME /e:/ and /ɛ:/ by using lexical set names? Clearly, using FLEECE for /e:/ and something else for /ɛ:/ would be inadequate on the grounds of FLEECE being something rather different now than what it was in ME in both quality as well as the number of lexical items it contains at either stage. The alternative would be to invent new lexical set names which display the insight that two vowels were different in ME, but have collapsed in EModE. For example, we might call ME /e:/ MEET and ME /ɛ:/ MEAT. Such an approach, however, would fail to capture the continuity between the two ME vowels and their modern reflex (apart from introducing oddities in contexts where the spelling is not apparent).

Throughout this thesis, I will therefore use IPA notation for designating pre-modern vowel phonemes, and lexical set names for contemporary vowels.



## 2.

## Methodology

## 2.1 Introduction

This chapter provides information on the speakers selected for the analysis (section 2.2), the variables (2.3) and the method of analysis (2.4). Section 2.5 specifies some terminology and abbreviatory conventions which will be used throughout this thesis.

In addition, section 2.6 discusses the question of vowel normalisation in some detail, as that has a bearing both on questions specific to this thesis (i.e. how raw data has been processed) as well as general points regarding the problematic status of vowel normalisation in this type of work.

## 2.2 The speakers

30 speakers from the intermediate archive (henceforth IA) held at the University of Canterbury were chosen for the analysis. The IA consists of speakers born between the late 19th century and the 1930s. This archive therefore complements two other corpora, the Mobile Unit corpus consisting of speakers born between the 1850s and the 1890s (cf. Gordon et al. 2004) and the Canterbury Corpus. The IA mainly consists of tape-recorded interviews carried out within the framework of two postgraduate projects on local history. The interviewers are former students of history at the University of Canterbury (Rosemary Goodyear and Sandra Quick). The interviews vary in duration between 10 minutes and 1 hour. The speech material from which the data analysed here is drawn is running speech (rather than word list material, which is also available on some of the recordings), the topics include the interviewees' local history, family background, descent, and childhood. All of the speakers analysed here were born in NZ. The sample (as well as the IA in general) is heavily biased in favour of speakers from the South Island, especially Otago. The speakers were chosen according to the following criteria:

- (a) An approximately even distribution across the overall timeframe covered by the IA (i.e. both 'late' and 'early' IA speakers)
- (b) A clustered distribution within that timeframe, i.e. only speakers who were born between 1890 and 1905, 1910-1920, after 1924. This was done in order to allow for meaningful subgrouping within the overall sample without arbitrary breaks along an even age continuum.
- (c) An approximately even number of male and female speakers.

Socio-economic factors were not carefully controlled for mainly because the overall degree of social stratification was small in the IA, especially as far as female speakers are concerned. That is, it was not possible to control for socio-economic status both across and within the subsamples.

The speaker sample was broken down into six subsamples, which will henceforth be referred to as early/medium/late males/females, respectively. Table 2.1 below shows the speakers' birthdate as well as the subgrouping.

<i>Group</i>	<i>Speaker</i>	<i>Birthdate</i>
Early males	C.N.	1902
	J.M.	1900
	T.McC.	1903
	N.C.	1905
Early Females	L.A.	1897
	D.H	1896
	J.B.	1898
	J.McL	1901
	V.H.	1902
	S.S.	1902
	N.A.	1903
Medium males	R.W.	1914
	Er.R.	1919
	B.G.	1922
	E.T.	1915
	J.J.	1922
Medium Females	M.H.	1915
	M.S.	1915
	I.R.	1918
	M.G.	1921
	P.G.	1921
Late Males	J.W.	1928
	T.K.	1930
	E.L.	1931
	B.A.	1928
Late Females	J.C.	1924
	El.R.	1928
	P.L.	1931
	S.B.	1931
	P.H.	1936

Table 2.1 – Speakers analysed in the present thesis and their birthdates.

### 2.3 The variables

Three groups of vowel variables were analysed:

1. The short front vowels KIT, DRESS and TRAP.
2. The front centring diphthongs NEAR and SQUARE.
3. A group of lexical items which variably show [æ] or [a:] across dialects of English (the so-called ‘broad A’ words, cf. below).

These variables were chosen for both synchronic and diachronic reasons: Synchronically, their realisation defines the vowel system of New Zealand English against its genetically closest varieties, i.e. South African English, Australian English and RP. For details cf. chapters 3-6. Diachronically, it has been shown (Gordon et al. 2004, Trudgill 2004) that speakers covered by the Mobile Unit recordings do not show the contemporary system (with regard to these variables), and it has been hypothesised that the time frame covered by the IA is constitutive in the development of the current spoken standard of New Zealand English.

## 2.4 Acoustic analysis

5 series of measurements were obtained:

- (1) Series 1 obtained F1/F2 measurements of short front vowel tokens from all subjects. It was attempted to obtain at least 50 tokens for each short front vowel from each speaker, which was not possible in some cases. Therefore, token numbers vary between 30-100 for each speaker per vowel. Only stressed tokens were measured. Measurements were taken preferentially at a turning point in the F2 contour. If no such turning point could be identified, the token was measured at approximately mid-point. The PRAAT program for acoustic analysis of digitised data was used (Boersma & Weenink, [www.praat.org](http://www.praat.org)). The formant tracking as calculated by PRAAT was rechecked by visual inspection of the spectrograms. Pole numbers were adjusted in case of unsatisfactory fit. If no good fit was found, the measurement was taken at the point of maximum amplitude. The total data pool consists of 4600 measured tokens.
- (2) Series 2 obtained F1/F2 measurements of all monophthongs from all speakers. This was done in order to allow for relational statements regarding formant frequencies, which is required both to evaluate differences in absolute quality as well as to feed normalisation algorithms (cf. section 2.6 below). In this series of measurements, it was attempted to obtain 10 tokens per speaker/vowel. In the case of high back vowels FOOT and GOOSE, this was not possible for each speaker. A minimum of 5 tokens were obtained in these cases.

Measurement points were assigned as above. The same software was used. The data pool consists of 600 tokens.

- (3) Series 3 obtained a restricted pool of duration measurements for the short front vowels. For each speaker from the early male and the late female subsamples, approximately 10 tokens of each short front vowel in pre-plosive environments were measured. Only alveolar stops were considered in the case of DRESS and TRAP as well as pre-velar tokens for KIT (for the relevance of the pre-velar environment cf. chapter 3). For each speaker, an even number of tokens before voiced and voiceless stops were measured for a total of 324 tokens.
- (4) Series 4 obtained measurements of NEAR and SQUARE tokens from every speaker. The measurements were carried out with akustyk (by Bartłomiej Plichta, <http://bartus.org/akustyk/>), a plug-in for PRAAT. For each token, both a point measurement and an interval measurement were taken. The point measurement was taken at the centre location of the nucleus of each diphthong token. The criterion for this was either a turning point in the F2 contour or the point of maximum amplitude as determined by visual inspection of the oscillogram. Tokens of various degrees of stress were included, and auditorily coded for three levels of stress. Tokens that were reduced to schwa or clearly a token of NURSE were excluded, as were tokens that are etymologically NURSE, but were realised as SQUARE. For the interval measurements, a sampling rate of 10 ms was used. Both formant frequency (in Hz and bark) as well as duration measurements were taken. The data pool consists of 1092 point measurements and approximately 17000 interval measurements. In addition to the acoustic measurements, each token was auditorily assigned to either one of the two vowel phonemes.
- (5) Series 5 obtained point as well as interval measurements of all potential broad-A vowels (cf. chapter 1 and 5). Both pre-nasal and pre-fricative environments were considered. The measurement procedure was the same as in Series 4. This data pool consists of 448 point measurements and approximately 5900 interval measurements.

The pooled data was subsequently coded for two non-linguistic variables (age and speaker sex) as well as a number of linguistic ones. The coding categories differed depending on vowel class and will be discussed separately in chapters 3-6. Since most of this thesis will be concerned with group averages rather than individual speakers, it was attempted to find an appropriate normalisation procedure. Since normalisation constitutes a serious manipulation of raw data especially in a context of diachronically stratified sample, it deserves a broader discussion, which is the topic of section 2.5 below.

Unless otherwise stated, formant frequency values represent normalised values throughout the paper.

## 2.5 Normalisation

This section discusses a number of points regarding the transformation of raw data to normalised data in order to investigate sociolinguistic variation independent of physiological differences between speakers. Section 2.5.1 states a number of aims one wants to achieve by normalising data in a sociophonetic context. In section 2.5.2, a brief overview of the vast literature on normalisation will be given. Section 2.5.3 applies a number of common normalisation techniques to a range of data from the context of the current study. Finally, a number of caveats will be discussed in section 2.5.4.<sup>6</sup>

### 2.5.1 Why normalisation?

The basic problem regarding the representation of vowels in sociophonetic studies is that physiological properties of speakers interfere with our ability to document sound change over time as well as synchronic patterning, i.e. social stratification. This is due to the fact that the basic representation of vowel structure in sociolinguistics is that of two-dimensional formant space. That is, vowel positions are expressed by plotting their first formant frequency against

---

<sup>6</sup> Much of what follows anticipates some of the results presented later on in this thesis. However, I here restrict the background knowledge to rather uncontroversial and widely recognised points regarding the history of New Zealand English such as that ‘there was/is a shift in the short front vowels’. Most of the following chapters are concerned with questions of detail.

the second formant frequency and thereby relate differences between speakers to differences in vowel articulation.

However, it is well known that physiological differences between speakers introduce variance in precisely the same dimension (formant frequency), whereby speakers with shorter vocal tracts have higher formant frequency values, all other things being equal. For example, in a vowel change that involves raising of a vowel phoneme in articulatory space, this process would be overstated for speakers with shorter vocal tracts. Given that vocal tract size correlates solidly with speaker sex as well as age (where female speakers have shorter vocal tracts than males, and children have shorter ones than both males and females), an analysis based on raw data uncontrolled for sex and age might therefore overstate the innovativeness of females in a change that involves either fronting or lowering of a vowel phoneme, as these processes involve an increase of formant frequency values over time. By the same token, a backing/raising process would intrinsically appear as being led by male speakers<sup>7</sup>.

Peterson and Barney (1952) is a classical study showing the degree of variation around a mean for an English monophthong system (American English) that is introduced by plotting pooled raw data from males, females and children onto an F1/F2 diagram. With respect to the present study, consider figure 2.1 below, which plots raw F1/F2 averages of KIT and DRESS as produced by a male speaker and a female speaker from the Intermediate Archive (the latter one being about 30 years younger).

---

<sup>7</sup> This is of course simplistic in a vowel change that involves changes in both the open-close as well as the front-back dimension since it relates to absolute vowel position only. The set-up of vowels relative to each other is not affected by different speaker physiologies in the same way.

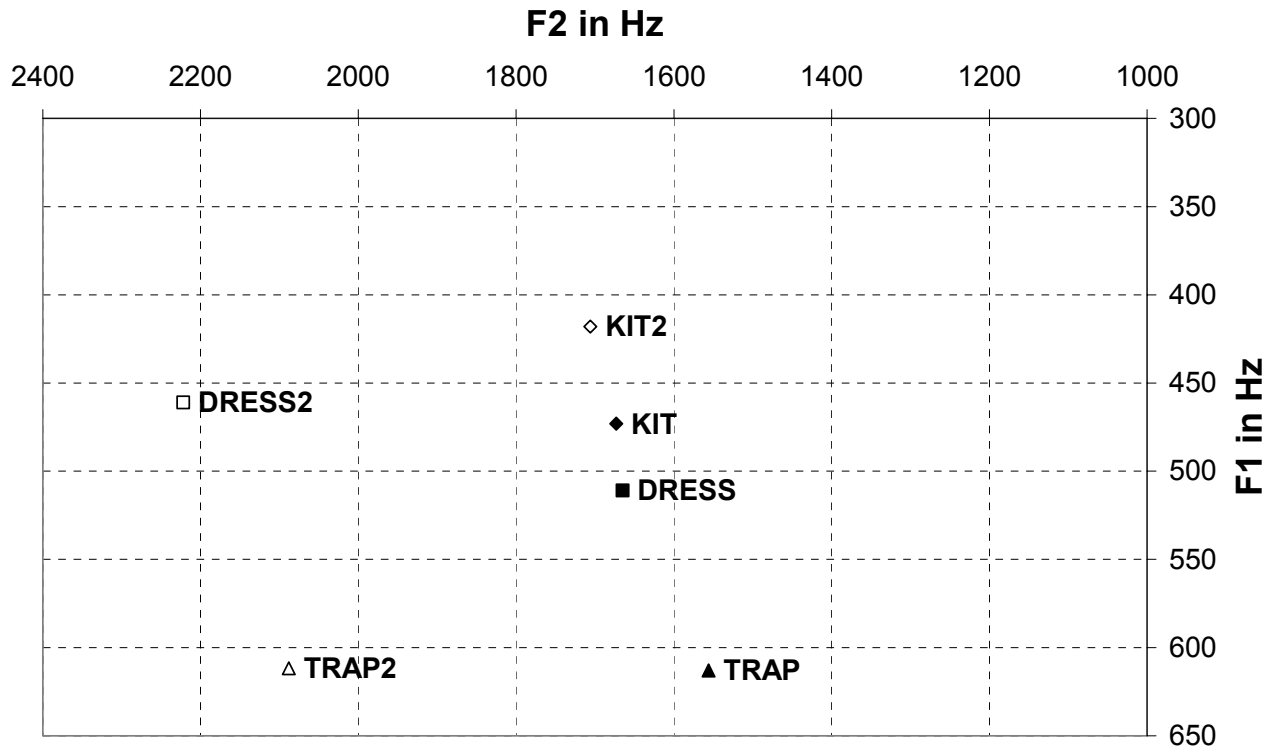


Figure 2.1 – Average raw frequency values of F1 against F2 of two speakers from the Intermediate Archive. The data points labelled ‘2’ are productions by the late female speaker, born about 30 years later than the male speaker.

This suggests that, all other things being equal, both speakers have similar realisations of KIT as well as vastly different realisations of DRESS (the female speaker showing a fronted vowel).<sup>8</sup> This observation is in contrast with both the auditory impression one gets by listening to the two speakers (although the female speaker does indeed have a fronter vowel for DRESS, she also has a much more central realisation of KIT than the male speaker) as well as with a priori expectations regarding the nature of the SFV shift in NZE, where we would expect younger female speakers to show more innovative realisations of both vowels. However, this also introduces a slight risk of circularity in that one might be tempted to presuppose the phenomenon under study as a control mechanism to check for the validity of normalised representations (but see section 2.6.4 below for a discussion). Therefore,

<sup>8</sup> But cf. footnote 1 above: Whereas absolute formant positions may be affected by physiological effects, the set-up of the three vowels relative to each other wouldn’t. Even with raw data as plotted in representations such as fig. 2.1, we can therefore make statements such as ‘the late female speaker has a more central vowel KIT than the early male speaker’.



normalisation should ideally factor out physiological differences and at the same time maintain differences that are due to sound change and its apparent-time representations.

Apart from being an important tool for representing vowels in sociophonetic research, a number of researchers (cf. Neary 1978, 1989; Syrdal & Gopal 1986) have introduced the notion of normalisation as reflecting a perceptual process whereby listeners normalise speech in order to resolve ambiguous acoustic information. That is, some kind of transformation of the raw input signal is assumed to occur in speech processing as to explain how speakers can, with great accuracy, perceive phonemic identity in the face of acoustic differences. Since the data presented here does not allow for making comments about speech perception, this point is irrelevant in the present study.

### *What normalisation should achieve*

When formant measurements of a number of tokens of various vowel phonemes are taken and plotted onto an F1/F2 diagram, this invariably introduces variance in both dimensions. As has been pointed out above, this variance can either be due to physiological difference between speakers or to different realisations of the vowel phonemes in question. We would therefore want a transformation which

- (a) Reduces overlap of distributions between different phonemes, all other things being equal

That is, if a group of speakers has a similar realisation of a vowel, we ideally want as large as possible a number of tokens to be closer to the average of that vowel than to that of an adjacent vowel. If all tokens are in ‘their own’ vowel space, overlap would be zero. In order to make a statement about the validity of a given normalisation procedure, we therefore want a measure of either (a) the distance of a token to its mean before and after a given normalisation procedure or (b) the number of tokens of a vowel A that are within the standard deviation of vowel B.

- (b) Maintains linguistically relevant distinctions, including distributional properties that are not due to anatomically induced formant frequency variation

An obvious thing that normalisation should seek to avoid is to collapse vowel phonemes, i.e. reducing overall variance around two means toward a common “centre of gravity”.

In the present context of a vowel shift involving three vowel phonemes, this can be checked for by calculating the distance ratios between vowel A (KIT), B (DRESS) and C (TRAP), which are expected to remain constant after the transformation.

In addition, sociophonetically relevant information should be maintained rather than normalised away. Coming back to the two speakers plotted in figure 2.1 above, we want the transformed values to reflect the auditory impression that speaker 2’s KIT phoneme is a central vowel, whereas speaker 1’s is a front vowel. In addition, we want the representation to capture the fact that speaker 2’s DRESS phoneme sounds rather like speaker 1’s KIT vowel. In order to check for the adequacy of a proposed normalisation procedure, we would need to resort to what we know about a given sound change beforehand. In this case, we can hazard the reasonable hypothesis that at some stage in the development of New Zealand English, the KIT vowel centralised and the DRESS vowel fronted. In addition, we would probably assume that this change was directional and that later speaker would tend to have more innovative realisations of both vowels.

I will therefore follow Hindle (1978) in assuming that in order to assess the resolving power of a normalisation procedure for sociophonetically pertinent factors, normalised formant frequency values should show tighter correlations with age than raw values. In addition, correlations with speaker sex should not be unidirectional across vowel phonemes in a speech community that is known to have undergone a change in more than one vowel and in different directions (i.e. centralisation of KIT and raising/fronting of DRESS). Finally, positive correlations between the relevant formant frequency values of a changing vowel and a stable vowel should only hold in the raw data.

## 2.5.2 A selection of formerly proposed normalisation procedures

A number of different normalisation techniques have been proposed over the last decades. These are generally divided into extrinsic (normalising a given vowel token by relating it to the realisation of other vowels within the sound system of the same individual) and intrinsic methods (which basically involves rescaling of hertz-values onto a perceptually more adequate scale such as ln (cf. Miller 1989), Bark (Zwicker & Terhardt 1980, and cf. Syrdal & Gopal 1986 for the role of bark in normalisation), mel (Stevens et. al (1937)) or ERB (cf. Glasberg & Moore 1990). Examples of extrinsic normalisation techniques include:

Gerstman (1968), who normalises data by rescaling raw formant frequency measurements using overall individual frequency minima and maxima, using the formula:

$$(1) F_{inorm} = \frac{F_i - F_{i\min}}{F_{i\max} - F_{i\min}}$$

Where  $i$  = the frequency of any formant, and  $F_{i\min}$  and  $F_{i\max}$  a given speakers minimum and maximum values for that formant, respectively. Normalised frequency values are therefore expressed as relative to a speaker's frequency extremes. In the present study, the average formant frequency of the vowel phoneme that shows the extreme values in the speech of a given individual were taken as the scaling values in order to minimise random effects of single tokens. For the remainder of this chapter, this procedure will be referred to as *gerst*.

Lobanov (1971), who derives z-scores by correcting a raw value with the centroid of all vowels and divides by the standard deviation of all vowels:

$$(2) F_{inorm} = \frac{F_i - \overline{F_i}}{SD_i}$$

Where  $F_i$  is the raw frequency of a given formant,  $\overline{F_i}$  is the average frequency of a formant  $i$  for a speaker, and  $SD_i$  is the standard deviation of that average. In the context of the present study,  $\overline{F_i}$  was obtained by measuring 10 tokens of each monophthong that is not a variable in

this study per speaker. In addition, 10 tokens of KIT, DRESS and TRAP were randomly selected. As opposed to `gerst`, this method normalises a given speakers formant frequency relative to that speaker’s centre of the vowel space rather than to the endpoints. We will refer to this procedure as `lob`.

Nearey (1978), who assumes constant ratios between vowel formant frequencies and normalises by transforming raw values to a log-scale and adds a speaker-dependent constant:

$$(3) \textit{Finorm} = F_i - \frac{(\overline{F_1} + \overline{F_2})}{2}$$

Where  $F_i$  is the log-transformed raw measure of a given formant, and  $\overline{F_1}$  and  $\overline{F_2}$  are the log-transformed averages of formants 1 and 2, respectively. A normalised value is therefore expressed as the distance of a token to a speaker dependent constant, namely that speaker’s overall ratio of the first two formants<sup>9</sup>. In the present study,  $\overline{F_1}$  and  $\overline{F_2}$  have been obtained by calculating the averages of the logarithmised 10 tokens of each monophthong per speaker mentioned in the context of `lob` above. This method will be referred to as `neary`.

Other transforms include that of uniform scaling based on higher formants (Nordstroem and Lindblom 1975) and Miller’s (1989) formant ratio normalisation. However, since the latter two require data on variables that were not measured in the context of the present study ( $F_3$  and  $f_0$ ), they will not be taken into account here.

Whereas intrinsic rescaling (from hz to bark, ERB, et cetera) has been shown to more accurately reflect vowel representations in speech perception compared to hertz scales, its usefulness in making sociophonetic claims is negligible (cf. Adank 2003).

A number of authors have compared various normalisation techniques to different data sets, and have concluded that as far as factoring out physiological differences and maintaining sociolinguistic ones is concerned, the procedures proposed by Lobanov and

---

<sup>9</sup> Note that Nearey (1978) proposes various methods. This one has been selected since it is the one that is usually drawn on in other studies comparing normalisation strategies such as Disner (1980). For a comparison of this as well as Nearey’s individual log-mean method with other procedures cf. Adank (2003) and Adank et al. (2004).

Neary fare best (cf. Adank (2003) for an extensive analysis of Dutch vowels. cf. also Disner (1980) for differences in the effectiveness of various normalisation techniques depending on whether the same or different languages are being analysed). I will compare these two (as well as *gerst*) techniques in the next section.

### 2.5.3 Normalising Intermediate speakers

Recall from section 2.6.1 above that in order to evaluate the effectiveness of a normalisation technique in a sociophonetic framework, a number of conditions should hold in normalised data vis-à-vis raw data. First of all, normalised data would be expected to reduce overall scatter around the mean of a given phoneme. This reduction would be due to the fact that physiological differences have been filtered out that enhance frequency minima and maxima within a vowel distribution. In addition, age effects that are present in the raw data should be preserved in normalised data if they can be shown to hold differentially across vowel phonemes. That is, an age effect that is present in one vowel but not in another can be hypothesised to reflect vowel change over time<sup>10</sup> and is therefore relevant to a socio-phonetic study.

In the following discussion, three sets of data will be analysed. In section 4.1 , all tokens of the vowels DRESS and TRAP from 9 late speakers will be looked at. This allows for making statements about scatter reduction around phonemic means in a maximally controlled environment<sup>11</sup>. The raw data will be compared to the transformed values using the techniques outlined above. A number of measures will be proposed to evaluate the output of each normalisation algorithm compared to the raw data. First, I will analyse the average F1 and F2 values of a number of vowels of each of the 30 speakers analysed in the context of this thesis. The vowels selected are KIT, DRESS and TRAP, which figure in the NZE front vowel shift, and STRUT, which is set as a ‘stable’ control variable. A number of analyses will be carried out on this data: A classification and regression analysis for both formants as well as correlation

---

<sup>10</sup> Whereas if an age effect is found that holds in all vowels , we are in all likelihood faced with a physiological effect whereby older speakers tend to show reduced vowel spaces (The evidence for age effects, however, is rather unclear. Although it seems clear that elderly speakers show systematically different vowel articulations, the nature of these differences is not clear, cf. Rastatter et al. (1997), Linville and Fisher (1985), Jacques and Rastatter (1990)).

<sup>11</sup> That is, since all tokens are from late speakers, the effect of vowel change in inter-sex comparison should be minimal. In addition, the two vowels change in the same direction. Proximity of male and female means in one vowel should therefore imply roughly the same proximity in the other vowel.

tests correlating the three dependent variables just mentioned with speaker sex, age and mean raw F1.

### *Token spread and frequency averages*

Figure 2.2 below plots all tokens of DRESS and TRAP from the late group of speakers (4 males and 5 females). 2 (a) shows the raw data, (b) the `lob` data, (c) the `neary` data and (d) the `gerst` data.

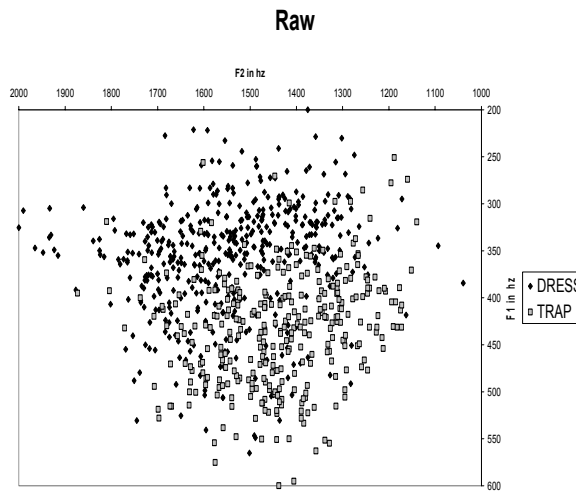


fig. 2.2 (a)

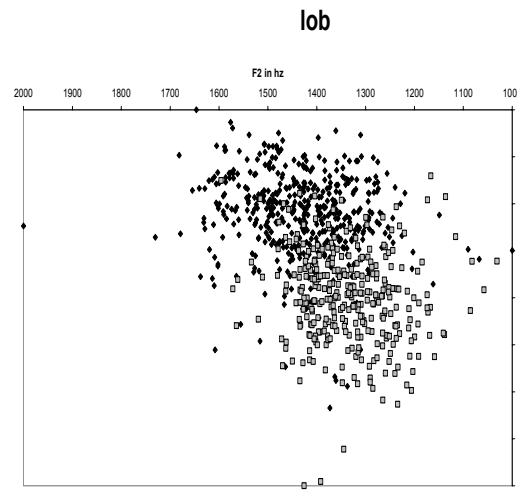


fig. 2.2 (b)

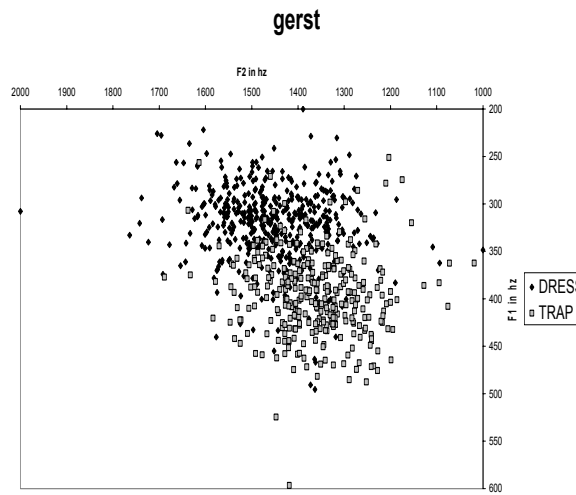


fig. 2.2 (c)

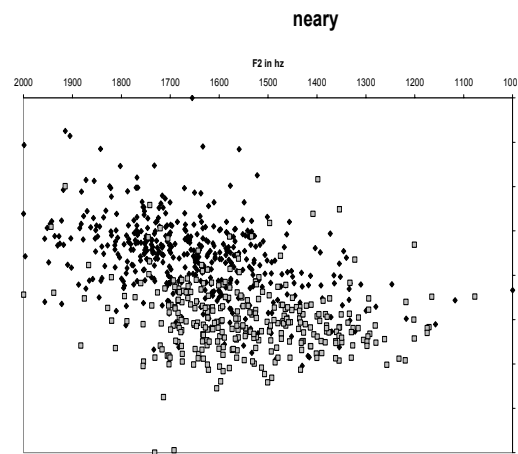


fig 2.2 (e)

Figure 2.2 – Token scatter of DRESS and TRAP in the speech of 9 late Intermediate speakers, 4 males and 5 females, using 3 different methods of normalisation, where 2 (a) shows the raw data, (b) is `lob`, (c) `gerst`, (d) `neary`.

As can be seen, all procedures reduce variance around the mean to some extent. What is less clear, however, is whether this reduction is indeed an adequate vehicle in analysing data in a sociophonetic context. We will consider a number of possible measures to assess this.

### *Scatter around the mean*

Table 2.2 below shows the mean values and their standard deviations for the vowels DRESS and TRAP in the late sample. The normalised data has been rescaled to match the minima and maxima of each formant in the raw data (i.e. the frequency value of the token with the highest/lowest  $F_n$  has the same value in both the raw data and the normalised data pools. The values are 200-600 for  $F_1$ , and 1000 – 2000 for  $F_2$ ).

		<b>DRESS</b>		<b>TRAP</b>	
		<b><i>F1</i></b>	<b><i>F2</i></b>	<b><i>F1</i></b>	<b><i>F2</i></b>
raw	<i>mean</i>	357.7	1540.6	433.6	1435.9
	<i>SD</i>	60.9	153.4	60.3	138.1
lob	<i>mean</i>	312.4	1432.5	400.9	1333.7
	<i>SD</i>	45.5	107.4	49.5	89.8
gerst	<i>mean</i>	321.6	1451.9	396.6	1363.6
	<i>SD</i>	40.7	109.5	45.2	98.5
neary	<i>mean</i>	380.2	1661.3	449.6	1542.5
	<i>SD</i>	43.4	157.9	39.5	152.3

Table 2.2 - Formant frequency averages and standard deviations of DRESS and TRAP in the late sample.

It can be seen that all methods reduce the standard deviation of both formants to some extent. That this is not purely the result of lower average values in the normalised data can be seen

by comparing the neary F1 data to the raw F1: The SDs of both vowels are reduced by similar margins while the normalised mean frequency average is higher in DRESS and lower in TRAP.

As for the F1 of DRESS, the reduction in standard deviation is highest in *gerst*, and lowest in *neary*. Both *gerst* and *lob* achieve similar reductions in F2; in *neary*, this is not the case. In the TRAP data, the reduction in F1 is greatest in *neary*. The opposite pattern holds for F2, where the SD is lowest in *lob*, and highest in *neary*.

### *Distances between vowel phonemes and male/female averages*

Recall that the main objective of a vowel normalisation procedure is to factor out variance in formant frequency due to physiological differences. Given that female speakers tend to have significantly shorter vocal tracts than males, we would expect a successful normalisation procedure to significantly reduce the distance between male and female formant frequency distributions. As a corollary of this, we would expect the overall distributions to be further apart from each other than is the case in the raw data. Table 2.3 below shows Hotelling-Lawley<sup>12</sup> trace scores for the phoneme averages of DRESS and TRAP as a whole as well as between the male and female averages of both vowels.

---

<sup>12</sup> The Hotelling-Lawley trace is a distance measure between token distributions, cf. Warren et al. (2004).



	DRESS - TRAP	male/female DRESS	male/female TRAP
raw	0.59 ( $p < .0001$ )	0.37 ( $p < .0001$ )	0.45 ( $p < .0001$ )
lob	0.99 ( $p < .0001$ )	0.01 $p > .2$	0.02 $p < .05$
neary	0.69 ( $p < .0001$ )	0.03 ( $p < .05$ )	0.1 ( $p < .0001$ )
gerst	0.84 ( $p < .0001$ )	0.001 ( $p > .6$ )	0.02 ( $p < .05$ )

Table 2.3 – Hotelling-Lawley trace scores between overall distributions of DRESS and TRAP as well as male and female distributions within one phoneme.

It can be seen that whereas in the raw data the distances between the male and female distributions of the same phoneme are only slightly smaller than that between the overall distributions of the two vowels, all normalisation methods have achieved the reduction of the latter to a fraction of the former. In the DRESS data, *lob* and *gerst* show more effective clustering of male and female tokens around the overall mean of a vowel distribution than *neary*, which maintains a difference between male and female speakers that is significant on the .05 level. This suggests that *gerst* and *lob* are more effective in token identification with respect to phoneme category, which is also supported by the data given in table 2.4 below. Table 2.4 shows the average Euclidean distance of formant frequency values from each of the two vowel distributions to the average value of both vowels (i.e., the first number in column 2/ row 2 shows the average distance of DRESS tokens to the overall average of DRESS).

	raw	lob	neary	gerst
DRESS to DRESS	141.45	97.5	107.9	96.46
DRESS to TRAP	187.6	159.4	157.7	146.2
TRAP to TRAP	130.1	87.6	106.2	90.4
TRAP to DRESS	180.3	151.5	153.9	142.1

Table 2.4 – Average Euclidean distances between all tokens of DRESS and TRAP to their own mean and the mean of the respective other vowel distribution.

A powerful normalisation procedure should yield lower distance measures of tokens to their own mean (i.e. distance of DRESS tokens to the average of DRESS) and higher ones to the other mean (i.e. DRESS tokens to the average of TRAP) compared to the raw data. This can be expressed as the ratio between the average token distance to the mean of the distribution it belongs to and the neighbouring distribution. These ratios are given in table 2.5.

	DRESS	TRAP
raw	1.32	1.39
lob	1.63	1.73
neary	1.46	1.45
gerst	1.52	1.57

Table 2.5 – Ratios of average Euclidean distances of DRESS and TRAP tokens, calculated as:

$d_i/d_j$ , where  $d_i$  is the average distance of tokens to the other phoneme average (e.g. DRESS tokens to the average of TRAP) and  $d_j$  is the average distance of these same tokens to their own mean (DRESS tokens to the average of DRESS).

It can be seen that all procedures increase the ratio of distances of tokens of a vowel  $i$  to the average of  $j$  and the average of  $i$  compared to the raw data. In both vowels, lob achieves the highest token distance ratios, followed by neary.

From the above considerations regarding overall scatter reduction and Euclidean distances, we can assume that all three extrinsic normalisation procedures improve the resolution of raw data into subdistributions around the average of a phoneme and reduce scatter that is due to differences between the sexes. In addition, we have seen that although neary reduces scatter around the mean best in absolute terms, this comes at the price of a

concomitant clustering of the overall data regardless of vowel identity. Therefore, *gerst* and *lob* are more effective in separating distributions of different phonemes and factoring out sex differences. However, since it can reasonably be hypothesised that sex differences are to some degree sociophonetically important (cf. the oft-quoted role of women as innovators in sound change ‘from below’, cf. Labov 2001) in a sample that is suspected to show sound change, a more thorough analysis is required.

Therefore, the next section will offer a somewhat more refined analysis of normalisation procedures in terms of interactions of factors such as sex, age and a number of intrinsic variables.

### *Modelling speaker averages*

The data considered in this section are the average formant frequency values of the vowels KIT, DRESS, TRAP, and STRUT of each speaker in the sample analysed in the present study. Whereas the former three vowels are variables under study in this thesis, the latter one can be regarded as a control variable. Impressionistically, I had assumed the STRUT vowel to be rather stable over the Intermediate period on an auditory basis.

Two methods of analysis will be presented: CART non-parametric regression modelling and correlation tests between a number of variables.

### *CART*

CART (Classification and Regression Tree) is appropriate for continuous dependent variables such as vowel formant frequencies:

The construction of classification trees is essentially a type of variable selection. [...] Classification trees are an attractive method of data exploration because they can handle interactions between variables automatically. They also have the advantage of being completely non-parametric. No assumptions are made about the underlying distribution of the data. These features make them less powerful for detecting patterns in the data, but fairly reliable in terms of the patterns found. Classification trees do assume that the effect being

modelled is organised into discrete factors. An analogous class of models, regression trees, deals with continuous data. [...] A classification tree begins with the data to be analysed and then attempts to split it into two groups [...]. Ideal splits minimise variation within categories and maximise variation across categories. (Mendoza-Denton et al. 2003, p. 128-29).

For the mathematical foundations of CART, see Breiman et al. (1984). The important thing with respect to the following exposition of internal factors involved in the NZE SFV shift are that divisions that CART finds are always binary at any node down the hierarchy of the overall set of divisions found in the data set, and that interacting categories are readily identified as successively branching nodes.

Figure 2.3 and 2.4 below show the CART representation of the data in terms of the variables age, sex, F1/F2 and vowel phoneme. Figure 2.3 shows the partitioning of the F1 data, Figure 2.4 that of the F2 data.

Note that here as well as throughout the paper, the length of any branch is unrelated to its significance. The representation is as follows: The variable along which a given data pool branches is given in bold. The subsamples which the data is split up into are indicated next to the branches. On terminal branches, the mean of the dependent variable is indicated (in italics). For each split, the branch with the higher value appears on the right hand side. For example, figure 2.3 (b) has to be read as follows: The most significant split in the data pool is found in the variable **vowel**, where one subsample comprises the KIT and DRESS data (with lower F1s), the other one comprising TRAP and STRUT (with high F1s). No further branching is found in the former pool, and the average F1 for that data is *413.4 hz*. The TRAP/STRUT branch is further split up along the variable **vowel**, with TRAP showing a lower F1 mean than STRUT.

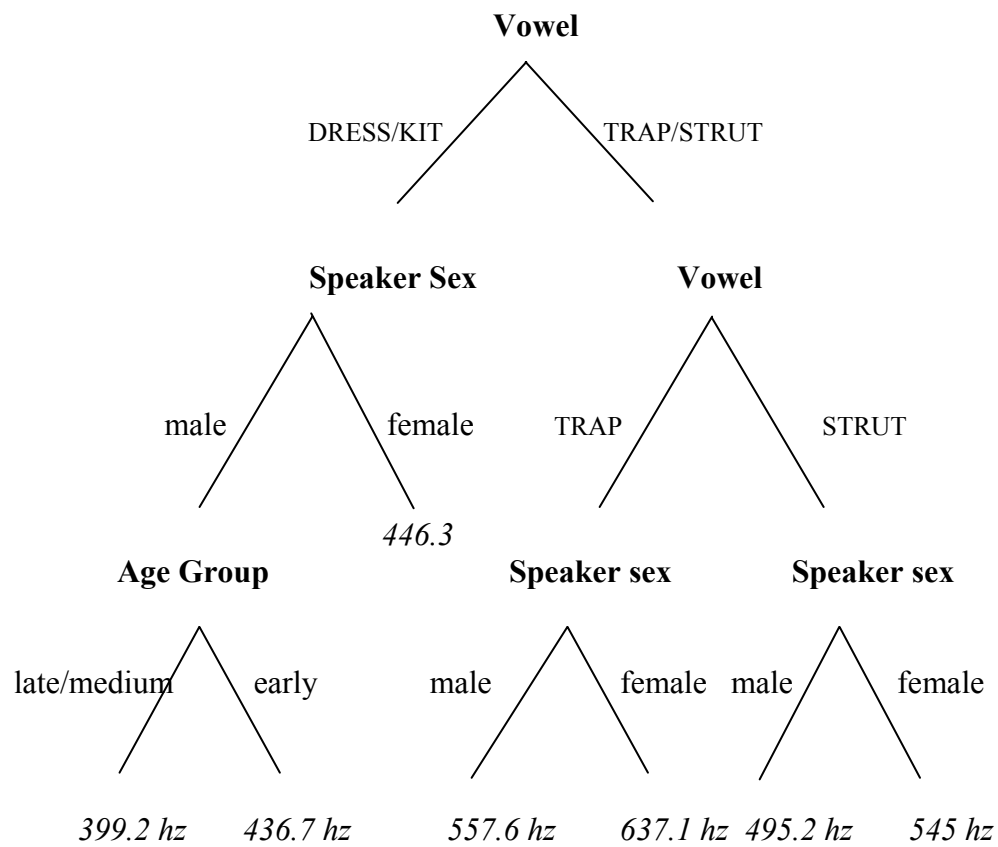


fig. 2.3 (a) – Raw data

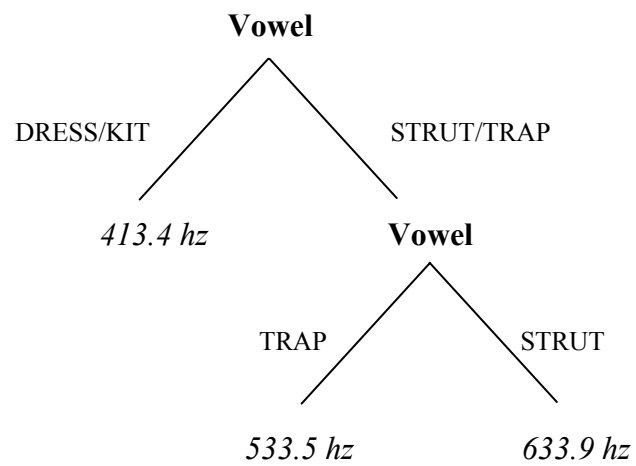


fig. 2.3 (b) - 1ob

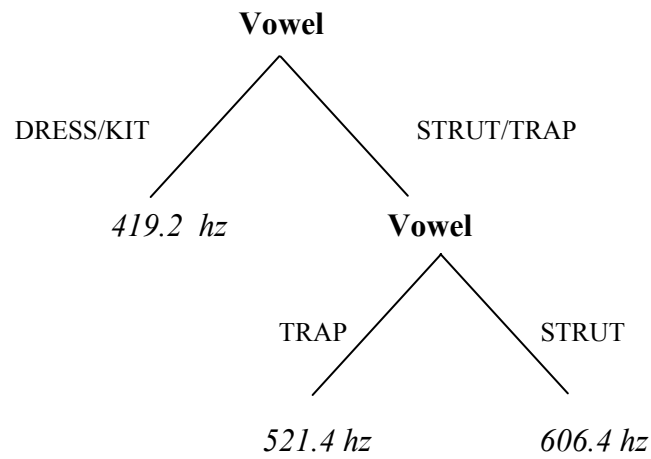


fig. 2.3 (c) - gerst

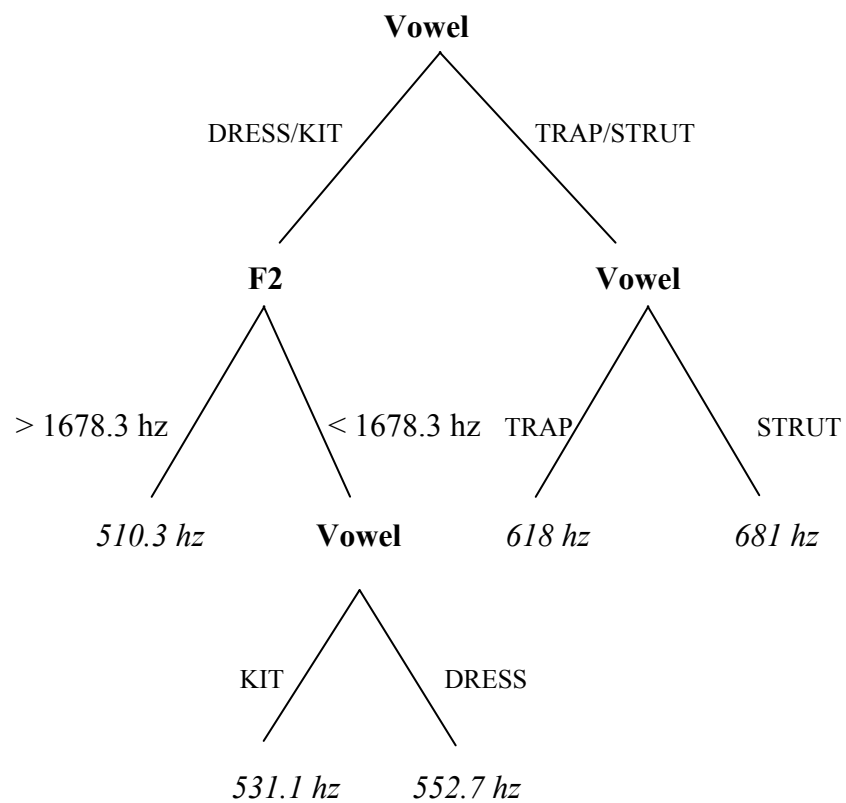


fig. 2.3 (d) - neary

Figure 2.3 – CART representations of pooled F1 data.

As for the F1 data presented in figure 2.3, the raw data is resolved first for phoneme, where DRESS and KIT (with low F1) contrast with TRAP and STRUT. This suggests that on the first branching point, the raw data indeed successfully recognises phonemic categories. The situation is more complex further down the tree: Within the DRESS/KIT node, speaker sex is the main predictor of F1 in the raw data. In addition, an age distinction is found within the male branch where late and medium speakers have lower F1 values than early ones. Whereas this is an expected outcome for the DRESS data, it is unclear why this should come about in a data pool that comprises both KIT and DRESS. As for TRAP and STRUT, the terminal nodes show a contrast in the variable sex, where female speakers show higher F1 values than males. On the whole, the raw data suggests that physiological factors play a major role in determining F1 values. The obvious alternative assumption would be to suggest that males lead lowering movements of all four vowels. This, however, seems to be ruled out by the failure of the variable *age* to show up in any of the low vowels within the overall male sample (i.e. TRAP and STRUT).

Both the *lob* and the *gerst* data as plotted in 2.3 (b) and (c) respectively resolve the data into vowel identity only, with no further subdivisions. On the one hand, this suggests that if the sex differences we observed above are mainly due to physiological factors, these have been filtered out by both procedures. However, the *lob* and *gerst*-normalised pooled data seems to suggest overnormalisation since no further variable appears in the analysis. What seems to be required here is an analysis of data from one vowel only. Fig. 2.4 plots the DRESS F1 data for both procedures.

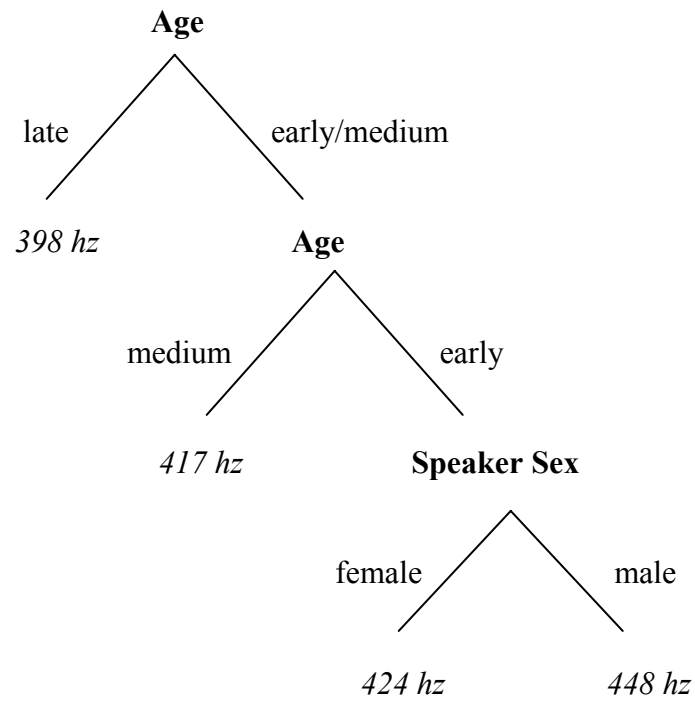


Figure 2.4 (a)

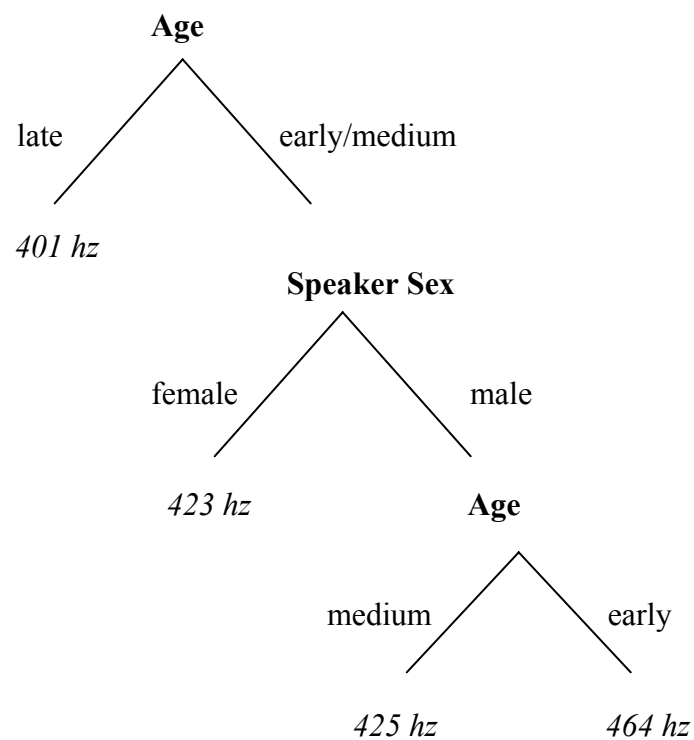


Figure 2.4 (b)

Figure 2.4 – CART representations of normalised formant frequency data in DRESS. 2.4 (a) shows 1ob, 2.4 (b) shows gerst.



Figure 2.4 shows that when data from individual vowel distributions are analysed, both `lob` and `gerst` recognise age and speaker sex differences, i.e. such patterns are not filtered out completely.

Finally, the `neary` data shown in fig. 2.3 (d) shows a slightly different pattern in that it splits the DRESS/KIT data into two groups distinguished by F2, where tokens of higher F2 show lower F1, and vice versa. It therefore makes a prediction that is contrary to what one would expect on the basis of physiological formant frequency dependencies (and may have captured a socio-linguistic pattern of variation that both `gerst` and `lob` miss). The patterning in the TRAP/STRUT data reflects that found in the other normalisation procedures.

## *F2*

Fig 2.5 shows the data for the second formant frequency values as modelled by CART. The raw data shown in 2.5 (a) suggest a primary split between male and female speakers, where female speakers show higher F2 values than males. Within the male branch, the data is further resolved into vowel identity, where STRUT contrasts with KIT/DRESS/TRAP. Interestingly, the subgrouping is slightly different within the female branch, where KIT falls into the STRUT branch, which is plausible assuming that female speakers lead a change that brings about lower (i.e. more STRUT-like) F2 values. In all branches except the one that includes only STRUT vowels, age differences are found at terminal nodes which are generally in the expected direction with the exception of a somewhat puzzling grouping under the rightmost node, where medium female speakers show lower F2 values than do late and early ones.

The normalised data (2.5 (b) – (d)) make rather similar predictions: The topmost division is within the variable vowel identity, where STRUT is split off KIT/DRESS and TRAP in all procedures. No further subdivision is found within the STRUT branch in any of them. In addition, all three normalisation methods split off DRESS from KIT and TRAP (the latter showing lower F2 values). Within DRESS, all three procedures indicate a negative correlation between F1 and F2. The only difference between the three is within the KIT / TRAP branch,

where `lob` and `gerst` state an age difference (figs. 2.5 (b) – (c)) with early speakers showing higher F2s in KIT and TRAP than late and medium ones, whereas `neary` (2.5 (d)) subdivides further into lexical sets (TRAP showing higher F2 than KIT).

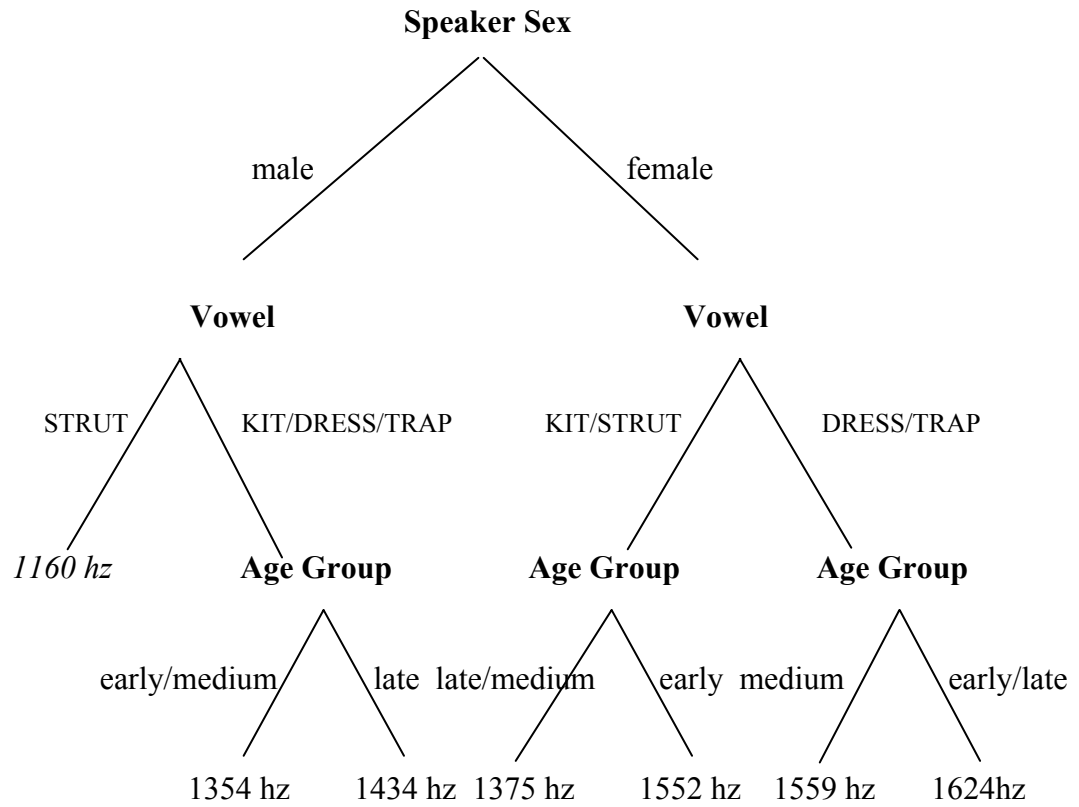


fig. 2.5 (a)

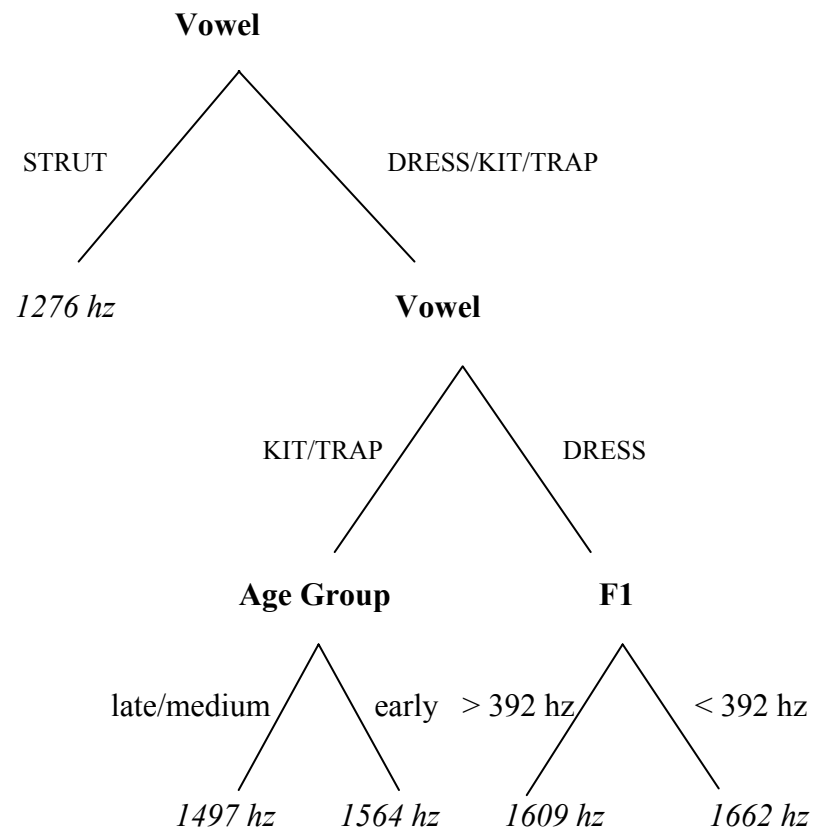


fig. 2.5 (b)

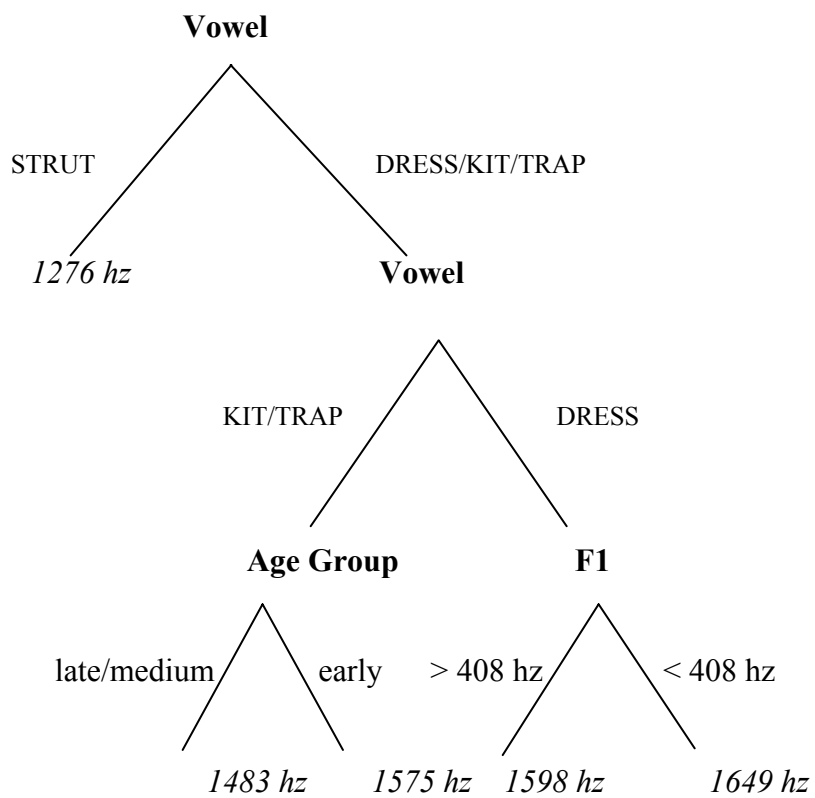


fig. 2.5 (c)

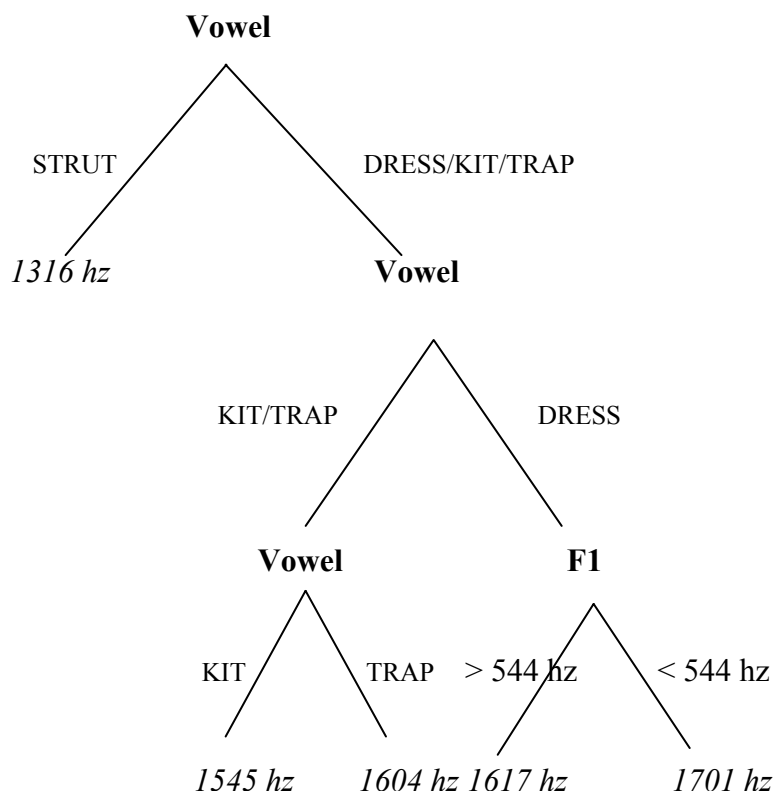


fig. 2.5 (d)

Figure 2.5 – CART representations of pooled F2 data.

### *Discussion (CART)*

Given the data as modelled above, we can conclude that all extrinsic normalisation procedures are to some extent successful in factoring out variation in the data that is in all likelihood due to differences in vocal tract length. That is, for socio-phonetic purposes any of the three procedures (*lob*, *neary*, *gerst*) is more adequate than raw data if statements are to be made about entire samples. We also observe that the three normalisation procedures make rather similar predictions regarding the partitioning of the data by using CART modelling, with minor differences at terminal nodes. It may, however, still be interesting to know which procedure is ‘correct’ with regard to these minor differences. *lob* and *gerst* essentially claim that in both KIT and TRAP, early speakers show higher F2 values than later ones. Given that we know that KIT underwent centralisation while TRAP underwent fronting

and raising, we have reasons to suspect that the pattern observed in KIT overrides that in TRAP, which might be due to either KIT tokens outnumbering TRAP tokens, or KIT centralisation being of a much higher degree than TRAP-fronting, or a combination of both. This may be assessed on the basis of the raw data, considering male speakers and female speakers separately. The relevant data is given in table 2.6.

		<b>KIT</b>		<b>TRAP</b>		<b>STRUT</b>	
		<b>F2 in hz</b>	<b>n of tokens</b>	<b>F2 in hz</b>	<b>n of tokens</b>	<b>F2 in hz</b>	<b>n of tokens</b>
<b>male</b>	<i>early</i>	1816	166	1703	186	1406	40
	<i>medium</i>	1609	314	1766	243	1355	50
	<i>late</i>	1801	200	1852	155	1429	40
<b>female</b>	<i>early</i>	2186	373	2198	338	1679	70
	<i>medium</i>	1819	295	2042	178	1534	50
	<i>late</i>	1866	209	2123	164	1517	50
<b>total</b>			<i>1557</i>		1264		300

Table 2.6 – F2 in hz and token numbers of KIT, TRAP and STRUT in all six subsamples. The formant frequency data is not normalised.

The raw data suggests a combination of both factors: There are more KIT tokens than TRAP tokens, and it is in KIT where late and medium speakers show lower F2s than early speakers. The subgrouping of late/medium vs. late is accounted for by the higher token numbers in the female sample, where this patterning is clearer in the KIT data (in addition, the high F2 value in the late male sample is to some extent carried by one rather aberrant speaker, who shows a rather restricted vowel space around a high mean in both formants. This peculiar pattern disappears after normalisation). Although the fronting of TRAP is reflected in the raw data on the whole, there is one exception: The early females show a much higher average F2 than any other group. In addition, the largest number of tokens is from this subsample, which in combination might neutralise the fronting effect in TRAP. In addition, the STRUT F2 average in this group is appreciably higher than in the other female sub-samples, which may be

indicative of systematically higher F2 values within that sample as a whole. The question is whether both *lob* and *gerst* have missed this effect. The data is inconclusive: If we regress F2 onto *age* and *speaker sex* for TRAP only, all three procedures find an effect whereby female speakers show higher F2s than male speakers (which is not surprising). However, *lob* and *neary* find an age group effect within the male branch only, whereas *gerst* finds one in the female branch (but not in the male one). We would therefore have to conclude that in order to find sociolinguistic patterns in normalised data, one cannot restrict the analysis to global data pools (such as a whole series of front vowels).

### *Correlations*

If it can reasonably be claimed that the overall sample shows variation over time in the realisation of one or more vowel phoneme but not in others, this should be reflected in correlations between the dimensions of change under hypothesis vis-à-vis ‘natural’ correlations that should hold between formant frequency averages and vocal tract length. That is, a successful normalisation should show tighter correlations between formant frequency and age (cf. Hindle 1978) than the raw data. Table 2.7 below shows correlation coefficients and significance levels for the following correlations: The average second formant frequency of KIT, the average first formant frequency of DRESS and the average first formant frequency of STRUT correlated with *year of birth*, and the *mean (raw) F1*. All data points represent the average for one speaker.

			raw	lob	gerst	neary
<b>KIT F2</b>	y.o.b.	cor.coeff.	<b>-0.41</b>	<b>-0.6</b>	<b>-0.64</b>	<b>-0.37</b>
		p	<b>&lt;.05</b>	<b>&lt;.001</b>	<b>&lt;.001</b>	<b>&lt;.05</b>
	mean F1	cor.coeff.	<b>0.52</b>	0.08	0.34	-0.06
		p	<b>&lt;.01</b>	>.6	>.1	>.7
<b>DRESS F1</b>	y.o.b.	cor.coeff.	<b>-0.66</b>	<b>-0.66</b>	<b>-0.61</b>	<b>-0.62</b>
		p	<b>&lt;.001</b>	<b>&lt;.001</b>	<b>&lt;.001</b>	<b>&lt;.001</b>
	mean F1	cor.coeff.	<b>0.86</b>	~0	0.04	0.36
		p	<b>&lt;.0001</b>	>.98	>.8	>.05
<b>STRUT F1</b>	y.o.b.	cor.coeff.	-0.1	0.36	0.1	0.32
		p	>.5	>.05	>.6	>.05
	mean F1	cor.coeff.	<b>0.82</b>	-0.1	-0.15	0.17
		p	<b>&lt;.0001</b>	>.6	>.54	>.37

Table 2.7 – Spearman correlation coefficients and significance levels in raw data and normalised data. Year of birth (y.o.b.) and mean F1 (raw) are correlated with the variables in the left-hand column.

On the whole, we observe that all normalisation procedures eliminate the tight correlations between the formant frequency variables in the left-hand column and the mean raw F1. This correlation is very high in the raw F1 data, and somewhat less so in the F2 raw data of KIT. The reason for the somewhat looser correlation between the latter and the mean F1 is probably due to a number of factors including (1) the non-uniform correspondence in size between different cavities (Fant (1973) notes that large vocal tracts are not simply scaled-up versions of small vocal tracts. Rather, the size differential increases more in the back cavity than the front cavity); and (2) female speakers leading the centralisation of KIT (i.e. speakers with higher intrinsic formant frequency values lead a change in which F2 lowering is involved).

Within the ‘stable’ variable STRUT we find a trend toward a positive correlation between F1 and year of birth in the lob and the neary data, but not in gerst (nor in the raw data). Both lob and gerst find stronger negative correlations between age and F2 (in KIT) than neary and the raw data. That the negative correlation is present in the raw data seems to be indicative of the age effect being rather more predictive of centralisation compared to the leading role of females in this process (i.e. both younger speakers as well as females would be expected to show lower extrinsic (i.e. sociolinguistically determined) F2 values for KIT than their complementary groups, while only the age is sufficiently strong to be reflected in the raw data).

Finally, we find in the both the raw data as well as the normalised data strong negative correlations between age and the first formant frequency of DRESS. Again, this seems surprising since the major effect that runs counter to the negative correlation would be an intrinsic one (where female speakers show higher F1 values due to physiological factors), which would be expected to disappear in normalised data. Therefore, normalised data would be expected to show a tighter negative correlation between the two variables in question.

#### 2.5.4 Discussion

In the foregoing sections, a number of analyses on various subsets of the overall data from the Intermediate sample have been carried out in order to evaluate the output of a number of normalisation strategies vis-à-vis the raw data and to compare these procedures with each other. In section 4.1, DRESS and TRAP tokens from 9 speakers who are of similar age have been analysed in order to make statements as to the effectiveness of vowel normalisation procedures in reducing scatter around the mean of the phoneme distribution as well as increasing the distance between phonemes. It was shown that although all procedures reduce scatter around the mean, *lob* and *gerst* are more effective in producing a high ratio of distance between phoneme categories and speaker sex averages. In addition, we have seen that all extrinsic normalisation transforms increase the distance ratio of tokens to the average of their phoneme distribution and the respective ‘other’ phoneme average. Again, *lob* and *gerst* were slightly more effective than *neary*.

The formant frequency averages of KIT, DRESS, TRAP and STRUT of all speakers were subsequently modelled using a CART analysis as well as correlation tests. The CART analysis revealed that on the first branching point in the F1 data, both normalised and raw data identify phonemic identity as the main predictor of formant frequency. On the terminal nodes, however, extrinsic normalisation procedures show no partitioning of the data along the variable *speaker sex* (with the exception of *neary*, where there is an F2 effect in the KIT/DRESS branch). Given that these factors as they show up on the terminal nodes of the raw data correlate well with what we would expect by physiological effects on formant frequency, we can assume that the normalisation transforms have successfully removed these effects.



The F2 data, albeit more complex than the F1 data independent of the type of transform applied to it, points in the same direction: Whereas speaker sex effects are found in all branches in the raw data, this is not the case in the normalised data. However, there are appreciable differences between the extrinsic normalisation methods as to the predictors found on the terminal and ante-terminal nodes. Both *lob* and *gerst* find an age effect within the KIT/TRAP branch, where late and medium speakers have lower F2 averages than early ones. *neary*, on the other hand, finds no age effect and splits the group of central tokens up into phonemic categories. Within the DRESS data, all normalisation procedures find an F1 effect whereby tokens of lower F1 have a higher F2. The important thing to note here is that this F1 effect is exactly the opposite of what we would expect if this represented an instance of unsuccessful filtering out of vocal tract effects on formant frequency.

Subsequently, both raw and normalised data of three vowels (KIT/DRESS/STRUT) was correlated with year of birth and mean F1 of all vowels and for all speakers analysed in the current study. In a successfully normalised data pool we would expect tighter correlations (compared to raw) data with year of birth in variables that are hypothesised to undergo change over that sample, as well as looser correlations (or no correlations at all) in a stable variable. These expectations were, by and large, borne out: In the raw data, significant positive correlations were found between any of the dependent variables and mean F1, which suggests that physiological properties are the strongest predictor of formant frequency in raw data (see below for some problems regarding this assumption). These correlations are not present in the normalised data, although trends exist in *neary* (DRESS F1 and mean F1) and *gerst* (KIT F2 and mean F1). Tight correlations are found between F1 of DRESS and year of birth in all data pools regardless of normalisation. In addition, strong correlations between year of birth and F2 in KIT are found in *lob* and *gerst*, and slightly less so in the raw data and in *neary*.

On the whole it appears as if all normalisation procedures make rather similar predictions regarding the formant frequency distribution within the current sample. Which one should we choose? It might be worth going back to the raw data in assessing this question. What properties would the raw data have to show in order to compromise a given normalisation procedure? *lob* assumes that all speakers within a sample have a similar distribution of vowels around a mean. If this were not the case, as might be the case if

subsample *a* had higher vowels than subsample *b* across the board, the output F1 values of subsample *b* will be too high (since with lower F1 values, the numerator would be larger). This hampers the effectiveness of `lob` in between-language comparison (cf. Disner 1980) and can also be hypothesised to have an effect on cross-dialectal comparisons as well as comparisons of language states over time. For `gerst`, a similar case can be made, if the languages (or dialects, or temporal states) to be compared have undergone a process whereby vowel spaces ‘shrink’, since the scale would be different (and speakers with smaller vowel spaces would have higher output values. Note that this is only relevant if the shrinking is in the maximum values of a given formant, since the minimum value is subtracted in both the numerator and the denominator in `gerst`). With regard to `neary` the matter is somewhat more complex: Recall that the algorithm in the form applied here is a distance measure to the log-mean of the first two formants. Therefore, each formant to some extent acts as a ‘normaliser’ for the other one. Given the above mentioned hypothetical scenario of directional shunting of a vowel space in one dimension only, this would still affect `neary` transforms, albeit to a lesser degree than `lob`. Assuming categorically lower F1 values (due to raising of a number of vowels, but no lowering of others) in a subsample *a* compared to another subsample *b*, *a* would show systematically lower logmeans than *b*, the extent of which is determined by how many vowels are affected (assuming that data from all vowels are fed into the algorithm), and to what extent they are affected. With respect to the raw data analysed here, we would then have to ask (a) whether there is evidence for directional shunting in the overall data pool; (b) whether this affects both F1 and F2; (c) if this effect is present, is it due to only a number of vowels, or is it present in all vowels; and (d) does this affect the corner vowels which are the basis of the `gerst` calibration. Whether these questions can be successfully assessed on the basis of raw data is questionable (cf. the following section). However, one can start out by comparing raw data from male and female speakers separately and thereby eliminate one major non-linguistic predictor of formant frequency in raw data, i.e. vocal tract size that is due to speaker sex (although one would still be left with individual differences in vocal tract size). Table 2.8 shows the results of Wilcox tests applied to both the overall data pool as well as each of the monophthongs<sup>13</sup>.

---

<sup>13</sup> Note that the number of measurements taken for the three short front vowels KIT/DRESS/TRAP vastly exceeds that of the remaining monophthongs since the former are variables under study in the current thesis. For the overall data pool, 10 tokens of each of the three short front vowels were randomly selected.

			<i>early</i>	<i>early vs. medium</i>	<i>early vs. late</i>	<i>medium vs. late</i>
<i>males</i>	<b>F1</b>	all	542.3	<b>495.8</b> <b>W=152694</b> <b>p&lt;.0001</b>	<b>486.6</b> <b>W=93756</b> <b>p&lt;.0001</b>	W=94938 p>.2
		KIT	472.5	<b>437.2 hz</b> <b>W=1432</b> <b>p&lt;.001</b>	<b>421.9</b> <b>W=791</b> <b>p&lt;.05</b>	W=758 p>.9
		FLEECE	399	<b>358.4</b> <b>W=1520</b> <b>p&lt;.0001</b>	<b>348.1</b> <b>W=974</b> <b>p&lt;.0001</b>	W=864 p>.2
		DRESS	518.8	<b>441.1</b> <b>W=1779</b> <b>p&lt;.0001</b>	<b>414.3</b> <b>W=1162</b> <b>p&lt;.0001</b>	W=912 p>.1
		TRAP	611	<b>541.9</b> <b>W=1734</b> <b>p&lt;.0001</b>	<b>530.9</b> <b>W=1034</b> <b>p&lt;.0001</b>	W=863 p>.2
		STRUT	672	661.8 W=1147 p>.2	653.6 W=700 p>.2	W=771 p>.8
		START	702.6	<b>674.2</b> <b>W=1293</b> <b>p&lt;.05</b>	684.6 W=716 p>.1	W=684.4 p>.5
		LOT	650.4	<b>618.1</b> <b>W=1391</b> <b>p&lt;.01</b>	<b>601.7</b> <b>W=863</b> <b>p&lt;.01</b>	W=859 p>.2
		THOUGHT	513.6	<b>438.6</b> <b>W=1784</b> <b>p&lt;.0001</b>	<b>442.9</b> <b>W=1044</b> <b>p&lt;.0001</b>	W=624 p>.2
		FOOT	471.3	<b>441</b> <b>W=1427</b> <b>p&lt;.001</b>	<b>415</b> <b>W=986</b> <b>p&lt;.0001</b>	W=887 p>.1
		GOOSE	426.9	<b>395.7</b> <b>W=1425</b> <b>p&lt;.001</b>	<b>410.6</b> <b>W=743</b> <b>p&lt;.01</b>	W=641 p>.2
		NURSE	526.5	<b>445.4</b> <b>W=1858</b> <b>p&lt;.0001</b>	<b>429.3</b> <b>W=1121</b> <b>p&lt;.0001</b>	W=856 p>.2

Table 2.8 (a)

Table 2.8 – Wilcoxon test results and significance levels for all speakers from the Intermediate sample. The data pool comprises 10 tokens per vowel for each speaker. Correlations that reach significance values below 5% are given in bold, those that are below 10% are given in italics. In each row, the first value states the average formant frequency of the data given in that cell. The second/third entry indicate in the medium and late column the W-value and the significance in difference between the distribution in that cell and the reference distribution in the early sample. The rightmost entry gives the W-value and the significance in difference between the distribution of the medium sample and the distribution in the late sample for each vowel. 2.7 (a) and (c) shows correlations with F1 for the male and female speakers, respectively, whereas 2.7 (b) and (d) show correlations with F2 for each of the two sex cohorts.

			<i>early</i>	<i>early vs. medium</i>	<i>early vs. late</i>	<i>medium vs. late</i>
<i>males</i>	<b>F2</b>	all	1515	1498 W=124698 p>.3	1519 W=72340 p<.1	W=87561 p>.3
		KIT	1761	<b>1581</b> <b>W=1399</b> <b>p&lt;.01</b>	1750 W=594 p>.9	<b>W=420</b> <b>p&lt;.01</b>
		FLEECE	2112	2106 W=996 p>.9	2125 W=570 p>.7	W=714 p>.7
		DRESS	1800	1861 W=820 p>.1	<b>1929</b> <b>W=354</b> <b>p&lt;.01</b>	W=564 p<.1
		TRAP	1699	<b>1754</b> <b>W=762</b> <b>p&lt;.05</b>	<b>1793</b> <b>W=404</b> <b>p&lt;.05</b>	W=648 p>.3
		STRUT	1406	<b>1355</b> <b>W=1277</b> <b>p&lt;.05</b>	1354 W=755 p<.1	W=768 p>.8
		START	1439	1459 W=939 p>.6	1413 W=706 p>.2	<b>W=987</b> <b>p&lt;.05</b>
		LOT	1133	1111 W=1043 p>.7	<b>1057</b> <b>W=788</b> <b>p&lt;.05</b>	<b>W=984</b> <b>p&lt;.05</b>
		THOUGHT	862	911 W=901 p>.4	851 W=666 p>.4	W=889 p>.1
		FOOT	1301	<b>1223</b> <b>W=1314</b> <b>p&lt;.05</b>	1299 W=651 p>.5	W=564 p<.1
		GOOSE	1601	1544 W=1201 p>.1	1571 W=672 p>.3	W=713 p>.7
		NURSE	1546	1572 W=787 p<.1	1565 W=498 p>.2	W=778 p>.7

Table 2.8 (b)

			<i>early</i>	<i>early vs. medium</i>	<i>early vs. late</i>	<i>medium vs. late</i>
<i>females</i>	<b>F1</b>	all	600.9	<b>551.3</b> <b>W=254193</b> <b>p&lt;.0001</b>	<b>571.1</b> <b>W=240443</b> <b>p&lt;.0001</b>	<b>W=162731</b> <b>p&lt;.05</b>
		KIT	522	506.9 W=1928 p>.3	522.6 W=1626 p>.5	W=1045 p>.1
		FLEECE	444.2	<b>387.4</b> <b>W=2717</b> <b>p&lt;.0001</b>	<b>398.5</b> <b>W=2778</b> <b>p&lt;.001</b>	W=1121 p>.3
		DRESS	553.3	<b>483</b> <b>W=2629</b> <b>p&lt;.0001</b>	<b>485.5</b> <b>W=2647</b> <b>p&lt;.0001</b>	W=1306 p>.7
		TRAP	688.8	<b>619.3</b> <b>W=2701</b> <b>p&lt;.0001</b>	<b>627</b> <b>W=2473</b> <b>p&lt;.001</b>	W=1109 p>.6
		STRUT	791.5	<b>736.2</b> <b>W=2506</b> <b>p&lt;.0001</b>	792.9 W=1888 p>.4	<b>W=727</b> <b>p&lt;.001</b>
		START	838.7	<b>799.6</b> <b>W=2230</b> <b>p&lt;.05</b>	820.1 W=2004 p>.1	W=1060 p>.1
		LOT	707.1	<b>664</b> <b>W=2296</b> <b>p&gt;.01</b>	693.6 W=1886 p>.4	<b>W=889</b> <b>p&lt;.05</b>
		THOUGHT	530.1	<b>469.2</b> <b>W=2668</b> <b>p&lt;.0001</b>	<b>477.7</b> <b>W=2580</b> <b>p&lt;.0001</b>	W=1221 p>.3
		FOOT	505.4	482.2 W=2091 p<.1	505.2 W=1713 p>.8	W=979 p<.1
		GOOSE	482.9	<b>415.4</b> <b>W=2821</b> <b>p&lt;.0001</b>	<b>459.6</b> <b>W=2118</b> <b>p&lt;.01</b>	<b>W=709</b> <b>p&lt;.0001</b>
		NURSE	545.8	<b>501.5</b> <b>W=2442</b> <b>p&lt;.001</b>	<b>499.4</b> <b>W=2541</b> <b>p&lt;.0001</b>	W=1273 p>.1

Table 2.8 (c)

			<i>early</i>	<i>early vs. medium</i>	<i>early vs. late</i>	<i>medium vs. late</i>
<i>females</i>	<b>F2</b>	all	1793	<b>1686</b> <b>W=237903</b> <b>p&lt;.001</b>	1754 W=222696 p>.1	<b>W=141832</b> <b>p&lt;.1</b>
		KIT	2206	<b>1805</b> <b>W=3038</b> <b>p&lt;.0001</b>	<b>1886</b> <b>W=2741</b> <b>p&lt;.0001</b>	W=1465 p>.1
		FLEECE	2546	<b>2466</b> <b>W=2215</b> <b>p&lt;.05</b>	2320 W=1619 p>.4	<b>W=1731</b> <b>p&lt;.001</b>
		DRESS	2320	<b>2159</b> <b>W=2516</b> <b>p&lt;.0001</b>	2365 W=1619 p>.4	<b>W=657</b> <b>p&lt;.0001</b>
		TRAP	2153	<b>2028</b> <b>W=2510</b> <b>p&lt;.0001</b>	2183 W=1558 p>.3	<b>W=621</b> <b>p&lt;.0001</b>
		STRUT	1679	<b>1534</b> <b>W=2675</b> <b>p&lt;.0001</b>	<b>1517</b> <b>W=2649</b> <b>p&lt;.0001</b>	W=1317 p>.6
		START	1614	1689 W=1812 p>.7	1624 W=1672 p>.6	W=1156 p>.5
		LOT	1227	<i>1164</i> <i>W=2103</i> <i>p&lt;.1</i>	<b>1169</b> <b>W=2141</b> <b>p&lt;.05</b>	W=1224 p>.8
		THOUGHT	960	<b>888.4</b> <b>W=2409</b> <b>p&lt;.001</b>	938.4 W=1998 p>.1	<i>W=997</i> <i>p&lt;.1</i>
		FOOT	1330	1334 W=1695 p>.7	1351 W=1602 p>.4	W=1202 p>.7
		GOOSE	1837	<b>1757</b> <b>W=2214</b> <b>p&lt;.05</b>	<i>1776</i> <i>W=2109</i> <i>p&lt;.1</i>	W=1205 p>.7
		NURSE	1853	1819 W=1926 p>.3	1894 W=1514 p>.2	<b>W=875</b> <b>p&lt;.01</b>

Table 2.8 (d)

Table 2.8 shows that there are indeed differences in formant frequency in a number of vowels across subsamples. The question is whether these differences reflect sampling error, linguistic change, or non-linguistic factors. Table 2.8 (a) shows the results for F1 values across the three male subsamples. It seems clear that the major differences hold between the early and

the non-early samples, where F1s in the early sample are higher across the board than in the other ones. This difference is significant in all monophthongs except STRUT. None of the differences between the medium and the late sample are significant. The fact that this pattern is so prevalent suggests that the high F1 values in the early sample are probably due to non-linguistic factors, for two reasons. First of all, I am not aware of a case of raising of all vowels within less than a generation in any language. Secondly, and more importantly, the differences in F1 between the samples are of different magnitude across vowels. For example, DRESS has an average F1 that is 18% higher in the early sample than in the medium sample, whereas the difference in LOT is only about 5 %. In cases where no change in the F1 dimension have been reported in the literature, the difference seems to be in the order of about 10% (KIT/FLEECE/FOOT/GOOSE). An adjustment of 10 %, however, suggests that the open vowels (START/STRUT) have undergone some degree of opening (which in turn suggests that the assumption made in section 2.4.2 about STRUT as a stable vowel is presumably wrong). This renders *gerst* problematic in the F1 dimension. As for *lob*, the question is whether the lowering and raising actually cancel each other out. This seems to be the case: Adjusting the F1 values of the medium male samples toward a range that is between 5% and 15% higher renders the differences between the two data pools non-significant. Since this range includes the F1 differences for vowels for which no change in F1 has been reported in the literature (cf. above), *lob* would normalise appropriately (in the F1 dimension, and for these subsamples).

The picture in the female F1 data pool is rather less straightforward (table 2.8 (c)). Although the above pattern holds where early speakers have higher F1s in many vowels than medium speakers, this cannot solely be ascribed to the early speakers being deviant across the board. In some vowels (STRUT/LOT/GOOSE), significant differences hold between the medium and the late sample as well, while in the former two (as well as START), no significant difference is found between the late and the early sample. In addition, the major differences between the late and the early sample hold in vowels for which change in the F1 dimension has been reported for NZE (raising of DRESS/TRAP/THOUGHT, possibly rounding of NURSE). And if the differences between the early and the late sample reflect linguistically conditioned one, we are indeed faced with a unidirectional lowering of F1 values, which would hamper the effectiveness of the *lob* normalisation. *gerst* does not fare much better in this sample,

since the F1 distributions of FLEECE are significantly different across the early and the late sample, whereas those of START are not, which would result in different scales for each subsample.

As for the male F2 data (table 2.8 (b)), there is little indication of problematic data for either *lob* or *gerst*. The overall distributions are not statistically different, nor are those of FLEECE and THOUGHT, which demarcate the extremes in F2 for all speakers. In the female sample (table 2.8 (d)), we observe again a pattern where the medium group shows systematically lower formant frequency values than the early one in a number of vowels (KIT/FLEECE/DRESS/TRAP/STRUT/THOUGHT/GOOSE). Since this (a) reflects the pattern observed in the F1 data and (b) is in some cases contrary to what we would expect (ephemeral backing of DRESS and TRAP as a property of one generation seems unlikely), we have reasons to assume that these low F1 values are largely due to physiological effects. A comparison between the overall data pool also suggests that the changes cancel each other out, since the F2 distributions between the early and the late sample are not statistically significant. Again, *lob* can therefore be assumed to make valid predictions. The case is somewhat different for *gerst*, since the F2 values in FLEECE differ vastly across the subsamples. More specifically, the late speakers show a much lower F2 than early ones, which would result in a smaller overall scale for that sample.

As for the cross-formant normalisation, one would have to check whether the ratios between formants are indeed correlated. Table 2.9 below shows correlations between log-transformed F1 and F2 values for each speakers in the sample.



		<b>all</b>	<b>males</b>	<b>females</b>
FLEECE	rho	<b>0.38</b>	<i>-0.17</i>	-0.06
	p	<b>&lt;.05</b>	<i>&lt;.1</i>	<i>&gt;.3</i>
GOOSE	rho	<b>0.45</b>	-0.08	-0.02
	p	<b>&lt;.05</b>	<i>&gt;.3</i>	<i>&gt;.7</i>
FOOT	rho	0.17	0.01	0.01
	p	<i>&gt;.3</i>	<i>&gt;.8</i>	<i>&gt;.8</i>
THOUGHT	rho	<b>0.39</b>	0.06	<b>0.42</b>
	p	<b>&lt;.05</b>	<i>&gt;.4</i>	<b>&lt;.0001</b>
LOT	rho	<b>0.45</b>	<i>0.18</i>	<b>0.21</b>
	p	<b>&lt;.05</b>	<i>&lt;.1</i>	<b>&lt;.01</b>
START	rho	<b>0.48</b>	<i>0.17</i>	0.04
	p	<b>&lt;.01</b>	<i>&lt;.1</i>	<i>&gt;.5</i>
STRUT	rho	<b>0.69</b>	0.11	<b>0.16</b>
	p	<b>&lt;.0001</b>	<i>&gt;.2</i>	<b>&lt;.05</b>
NURSE	rho	<i>0.36</i>	<i>-0.18</i>	-0.1
	p	<i>&lt;.1</i>	<i>&lt;.1</i>	<i>&gt;.1</i>
KIT	rho	0.3	<b>-0.22</b>	<b>-0.16</b>
	p	<i>&gt;.1</i>	<b>&lt;.05</b>	<b>&lt;.05</b>
DRESS	rho	0.31	<b>-0.23</b>	<b>-0.22</b>
	p	<i>&gt;.1</i>	<b>&lt;.05</b>	<b>&lt;.01</b>
TRAP	rho	<b>0.65</b>	0.02	~0
	p	<b>&lt;.001</b>	<i>&gt;.8</i>	<i>&gt;.9</i>

Table 2.9 – Correlation test results between log-transformed F1/F2 and significance levels for all speakers from the Intermediate sample. The data pool comprises 10 tokens per vowel for each speaker. Correlations that reach significance values below 5% are given in bold, those that are below 10% are given in italics.

Positive correlations between log F1 and log F2 are found in most vowels (exceptions are KIT, FOOT and DRESS. The correlation in NURSE is a trend below the 10% level of significance). It seems clear that these correlations are carried by the male/female differences, since within each the male and the female sample, they do not hold (with the exception of THOUGHT and LOT in the female sample). In addition, we observe negative correlations for the KIT and the DRESS data, which indicates that linguistic factors can override physiological correlations. Therefore, caution seems warranted if we take a log transformed cross-formant all-vowel average as the basis of normalisation of vowel data.

### 2.6.5 A brief meta-discussion

Most of the above analysis is rather speculative, and one might state with reasonable justification that there cannot be absolute certainty whether normalised data actually normalises appropriately. This is for two reasons: First of all, there does not seem to be a straightforward way of independently assessing either the input or the output of a given normalisation procedure. That is, if formant frequency data is all one has, it is in principle impossible to check whether the input conditions of the algorithms are indeed met. This makes the notion of normalisation *circular*. For example, there is no way of checking whether the calibration vowels required by *gerst* remain stable over the time frame covered by some corpus. Resorting to auditory impressionistic assessments (cf. Disner 1980), or to maximal control of a corpus (cf. Adank et al. 2004) does not help either, since such approaches either deprive acoustic analysis of one of its greatest advantages (scientific objectivity, which becomes moot if it requires to be cross-checked by a much less reliable measure, i.e. auditory analysis), or are indeed circular themselves (if I maximally control a speech sample in terms of external variables such as geographical origin, social class, and the like, I still do not know whether such a sample is *linguistically* stable. This rather represents a leap of faith based on findings reported in the literature).

In addition, as Disner (1980) makes clear, and which should also be evident from the data presented in this thesis, the validity of any proposed normalisation procedure depends on the nature of the linguistic material that is subjected to it, since different algorithms require different input conditions. Again, this brings up the conundrum of having to know in advance of what one wants to eventually find out.

Is there no way out? In principle, there are three alternative solutions to presenting vowel formant data without having to resort to traditional types of normalisation. (1) One might refrain from presenting group data altogether. This, however, requires separate analyses of large numbers of individual speakers and may not be too insightful if properties of languages (rather than idiolects) are under investigation. One would also lose the possibility of carrying out large-scale statistical analysis of vowel formant data, since one would compare apples to oranges. (2) Vowel formant productions might be cross-checked with physiological (i.e. resonant cavity) measurements for each speaker. Whereas this would

probably be the most valid approach in scientific terms, it would also be unrealistic since it would require much more time to analyse one individual and would furthermore require technology that is not available in most linguistics departments. (3) Rather than using formant data plotted on F1/F2 charts, relationships between vowels might be expressed as distance measures, or more specifically, distance ratios. To give an example from the data analysed in this thesis: If two given speakers are hypothesised to show the (linguistically) same short front vowel space, they should show similar distance ratios between the three vowels (i.e. “the distance of DRESS to KIT is .4 of that of DRESS to TRAP”) regardless of absolute formant frequency values. One would of course lose the convenient mapping of acoustic data onto 2-dimensional charts, in addition to losing any type of anchor points in vowel space. On the positive side, such an approach might be able to readily incorporate other phonetic dimensions such as duration and diphthongisation. In the final chapter of this thesis I will articulate this idea more comprehensively.

For the purposes of the analytical chapters in the present thesis, however, I have used normalised data throughout in order to allow for general statements that can be conveniently illustrated using conventional F1/F2 group plots. In spite of the drawbacks and limits previously discussed, at least *gerst* and *lob* were shown to yield data that both reduces scatter that is presumably due to physiological factors while also retaining external factors, most importantly age. That is, the pathways of change over time do not seem to get swamped by normalisation. Of the two, *lob* seems to be a slightly more solid choice due to the larger number of control factors that enter into the algorithm (i.e. all vowels contribute to the scaling factor rather than just two in each formant frequency dimension, as is the case in *gerst*).

However, it has to be borne in mind that we have reasons to believe that the overall vowel system has shifted toward a closer configuration over the intermediate period, and that this would inherently bias the normalised data. More specifically, since the overall average is subtracted from the raw measurement in the numerator, a bias toward lower values would yield *lob*-transformed F1 values that are too high. Therefore, F1 values for later speakers may be overstated in the normalised data that is analysed in the following three chapters. With respect to the front vowel shift as well as the NEAR/SQUARE merger, this would imply that the data given for later speakers may be too conservative throughout. This, however, can

be argued to constitute ‘good bias’, i.e. bias against some of my claims in the following chapters, the most important of which are that there is both short front vowel raising (at least in DRESS and TRAP) as well as merger of NEAR and SQUARE as a result of raising of the lower vowel, and that both processes are well under way in the intermediate period.

#### 2.5.6 Conclusion

The above analysis suggest that overall, lob and gerst are more effective in carving out the sociophonetically relevant properties of this data set as age effects are preserved better than is the case in neary. Both Disner (1980) and Adank et al. (2004) favour either lob or neary as the most robust procedures, depending on the type of data analysed. Therefore, the combined evidence seems to make lob the best choice. The normalised formant frequency values in this thesis are derived from lob.

## 3.

## The Short Front Vowels - KIT/DRESS/TRAP

## 3.1 Introduction

This chapter presents the results of an acoustic analysis of the short front vowels (henceforth SFVs) in the speech of 30 speakers of Intermediate New Zealand English (henceforth NZE). It has frequently been pointed out that the realisation of these vowels in NZE is different from that in most other standard varieties of English, in that the KIT vowel has a more centralised realisation, whereas DRESS and TRAP are both raised and fronted (Bauer 1986, Wells 1982).

Section 3.2 gives some background information on the SFV shift in NZE, and a brief outline of earlier work on this subject. In section 3.3 I will present the overall patterns found in the analysis, and argue for a push-chain scenario and elliptical distributions in transitional vowel systems. Section 3.4 will be concerned with the results of a CART (classification and regression tree) analysis of a number of phonemic conditioning factors. Section 3.5 presents a number of correspondence tests to corroborate the hypotheses put forward in sections 3.4 and 3.5. Section 3.6 analyses the role of lexical frequency in the SFV shift. Duration will be analysed in section 3.7.

## 3.2 Background

## 3.2.1 Etymological considerations

The contemporary front vowel set of Standard English is the result of a number of significant changes in late Middle English/ Early Modern English. Although these mainly affect the long vowels, there are a number of implications relevant to the short vowel system (this will be

discussed in more detail in the final chapter of this thesis, I will give only a brief overview here).

### *The long vowels*

Late ME had a system of 4 non-back long vowels<sup>14</sup>, /i:/, /e:/, /ɛ:/ and /a:/. Between the 15<sup>th</sup> and the 17<sup>th</sup> century, a complex pattern of vowel movements occurred which fundamentally changed this system in both quality and systemic typology. In brief, the high vowel(s)<sup>15</sup> diphthongised, and non-high vowels raised by one slot. This development is commonly referred to as the Great Vowel Shift. Subsequently, the reflexes of ME /e:/ and /ɛ:/ merged in most varieties. The merged vowel is the current FLEECE vowel. In addition, raised ME /a:/ diphthongised to /eɪ/ (FACE), which results in a system with no non-high front monophthongs in contemporary Standard English (the /a:/ slot was later filled by a number of allophones of ME /a/ and /au/, cf. ‘palm’, ‘start’, ‘dance’, ‘half’. The details of this development, however, are somewhat complex and will be reviewed in chapter 6).

### *KIT*

The KIT vowel has remained relatively stable in most varieties since OE times. That is, most words that now contain the KIT vowel have had a short high front vowel in all stages of the history of English, and correspond to a similar element in cognate languages (cf. English *sit/wit*, German *sitzen/Witz*, i.e. ‘joke’). In addition, a number of transfers occurred: (a) OE /y/ unrounded, and words containing that vowel are now members of KIT (normally KIT words spelled with *u* such as ‘busy’, but also a number of *i*-spellings, cf. *midget* vs. German *Muecke* /mʏkə/), (b) a number of DRESS vowels preceding /ŋ/ have gone over to KIT (e.g. *England*), (c) shortening in trisyllables transferred /i:/-words in that environment to KIT (cf. *divinity* vs. *divine*), and (d) etymological KIT before velar fricatives lengthened to a long high vowel and diphthongised to /ai/, i.e. are now members of PRICE (cf. *light* vs German *Licht* /lɪçt/).

---

<sup>14</sup> The term ‘non-back’ is used here since ME /a:/ (presumably central) interacts with the front vowel system in the subsequent history of English.

<sup>15</sup> Plural *s* occurs in parentheses here since the GVS affected both front and back vowels, although the latter are not discussed in the present context.

## DRESS

As with KIT, DRESS has remained relatively stable since ME times. It derives from a non-high, non-low front vowel<sup>16</sup> [e] ~ [ɛ] and has been contributed to mainly by shortening of a number of ME long /ɛ:/ words (normally spelled with *ea* such as *head*, *dead*, *dread*). Earlier developments (late OE) add a number of words containing /e:/ via shortening before non-homorganic codas (hence the alternation in *keep/kept*) and transferred /e/ words to /ɛ:/ before homorganic clusters (cf. English *field* vs. German *Feld* /fɛlt/; the high vowel in English is a result of the GVS).

## TRAP

Unlike KIT and DRESS, TRAP is a relatively recent addition to the front vowel set of English. It derives from a ME low central short vowel; a quality that is preserved in most cognate languages (cf. English *cat/rat* vs. German *Katze/Ratte* /katsə/ /ratə/). The fronting to /æ/ occurred in late ME times (Lass 1990), and affected only the Southern varieties of English (and was hence assumed to have been a reactionary movement to the lowering of ME short /u/, i.e. the current STRUT vowel, on the grounds that the Northern varieties have neither a STRUT vowel, nor lowering thereof, nor fronted ME /a/). Furthermore, in some varieties with a fronted TRAP vowel, a lexical split occurred which transferred a number of words containing TRAP to a new lexical set, START (the details of this development, as well as its relative chronology, will be discussed in chapter 6)<sup>17</sup>.

## 3.2.2 SFVs in related varieties

Recently, the SFVs have been involved in a number of quality changes in various varieties of English. In Northern American English, they have been shown to form an interrelated chain

<sup>16</sup> There has been, however, considerable debate regarding when this vowel was [e], and when it became [ɛ], cf. Lass (1990), Trudgill (2004).

<sup>17</sup> The development of low vowels in English is strikingly repetitive as well as non-directional, i.e. involves fronting as well as backing: early OE /a/ was fronted to /æ/ in late OE, backed to /ɑ/ in early ME, fronted (in Southern varieties, cf. above) to /æ/ in late ME (with concurrent backing after /w/, cf. *cat* vs. *what*), and now undergoes re-backing in Southern British English. Cf. chapter 6 for some implications of this pattern for theories of directionality in vowel change.

shift (the Northern Cities Shift, cf. the extensive documentation in Labov (1994/2001)) whereby TRAP lengthens and rises, whereas KIT and DRESS lower. In the Southern Hemisphere Englishes, the non-high SFVs have been claimed to have undergone raising and fronting (cf. 2.3 below). KIT behaves somewhat less regular across the Southern Hemisphere varieties: Whereas in AusE it has raised/fronted to [i], it centralised to [ɜ] ~ [ə] in NZE. SAE has developed both qualities via allophonisation (cf. below).

### 3.2.3 The SFV shift in New Zealand English

The basic mechanism of the NZE front vowel shift can be described as a process that converts a system with three front vowels into one of two front vowels (DRESS and TRAP) and a central vowel (KIT). This can be sketched as in figure 3.1 below:

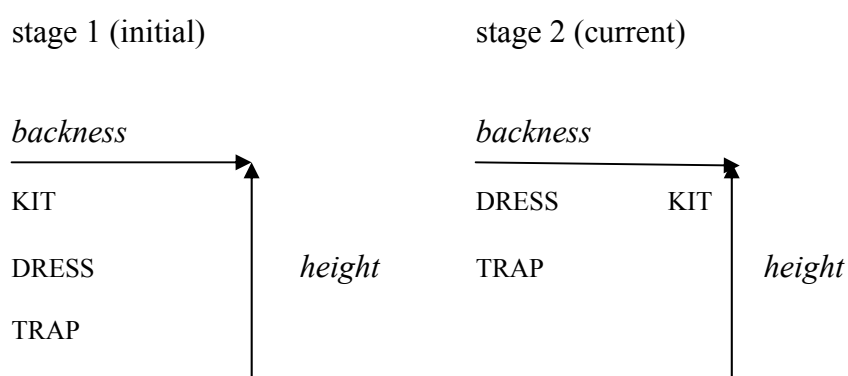


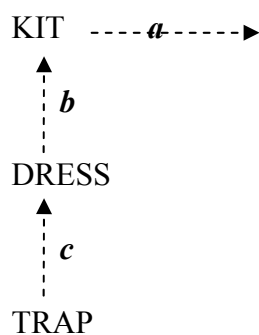
Figure 3.1 – A schematic representation of the New Zealand English SFV shift.

A representation such as the one shown in figure 3.1 leaves a number of questions open: First of all, it compares only an initial to a final stage of a dynamic process spanning over a number of generations and does not elucidate the properties, nor the mechanisms of the change. That is, it makes no assumptions regarding whether the changes in each of the elements involved in the process (i.e. the three vowel phonemes) are in any way related to each other. However, given that numerous examples of interaction in vowel systems have been demonstrated (at least since Luick 1914, and empirically since the quantitative research



carried out by Labov and others since the 1970s), and that the vowels involved in the restructuring of the NZE front vowel space are rather similar to each other on both etymological and phonetic grounds (etymologically they all derive from adjacent short front vowels, phonetically they have been assumed to be ‘short’, i.e. of lesser duration than comparable ‘non-short’ vowels such as FLEECE), the commonplace assumption is that the steps are interrelated and therefore form a chain-shift, a quality change in a number of adjacent segments of the same type under the maintenance of the original set of phonemic distinctions.

Therefore, a more adequate representation of the chain-shift would be as sketched in figure 3.2:



where *a* **because of** *b*, and *b* **because of** *c*

**or**

*b* **because of** *a*, and *c* **because of** *b*.

Figure 3.2 – A schematic representation of the NZE front vowel shift, where *a/b/c* represent the processes associated with each of the elements, i.e. *a* = centralisation of KIT, *b* raising of DRESS, *c* = raising of TRAP.

The statement sketched in figure 3.2 is rather more complex than a mere comparison between an initial and a final stage: It posits a chain of causal implications, that is, neither of the *caused* events would have happened without the other ones. We note further that a mere comparison between an initial and a final stage leaves open the question as to which processes initiated the overall process. Given the widely assumed causal mechanism of adjacent vowels interacting with each other (and only with each other, i.e. ‘long distance

processes' are generally not recognised as being evidenced by gradual vowel change, but may play a role in other types of change, cf. chapter 6/8), there appear to be three possibilities:

1.  $a > b > c$     2.  $c > b > a$     3.  $b > a/c$

where  $>$  = *precedes*

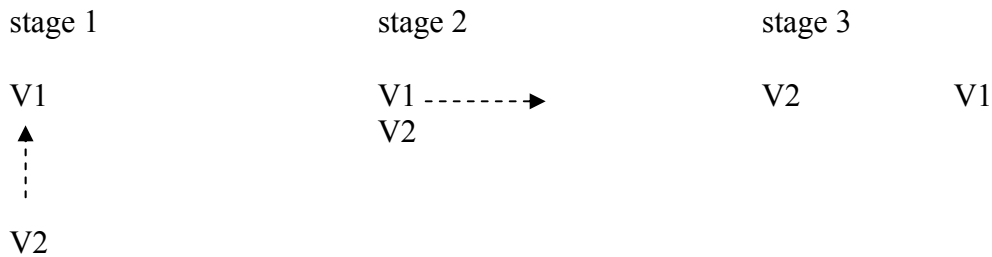
(1) states that the centralisation of KIT preceded the raising of DRESS, which in turn preceded the raising of TRAP. This type of chain-shift has been referred to as a 'pull-chain', and is evidenced by e.g. the later stages of the Great Vowel Shift, where ME /a:/ raises to /ɛ:/ after /ɛ:/ raised to /i:/. (2) states that the initiating process was the raising of TRAP, which led to the raising of DRESS, which in turn led to the centralisation of KIT. This type of process has been referred to as a 'push-chain'<sup>18</sup>. Finally, (3) assumes both a push- as well as a pull-relationship between the three vowels, in that DRESS may have raised first, which provided the incentive for KIT to centralise and for TRAP to raise. Such a pattern may have been the case in the Great Vowel Shift<sup>19</sup>, but has not been advocated in the case of the NZE front vowel shift, where earlier discussions revolved around arguing for either (1) or (2).

A further issue worth discussing is the relative timing of each of the sub-processes. That is, it is not clear *a priori* whether such a change comes about in small increments, and some caused process *b* commences only after some causing process *a*, or whether on some level, one implies the other. More specifically in the case of a push-chain, does the causer vowel encroach upon the space of the vowel it eventually pushes out, or do the two vowels remain approximately equidistant over the entire process? The two possibilities are sketched in figure 3.3:

<sup>18</sup> There are a number of theoretical problems associated with the notion of 'push-chain'. These will be discussed in chapter 6.

<sup>19</sup> On the account that the ME mid-vowels raised, which led to the diphthongisation of the high vowels /i:/ and /u:/ (cf. Lass 1990/Luick 1914 for an incisive argument in favour of this scenario. The traditional view is that of a pull chain, cf. Jespersen 1909, Stockwell 1972). The raising of the low vowels clearly postdates both of these processes.

### Scenario 1: Encroaching



### Scenario 2: Equidistance

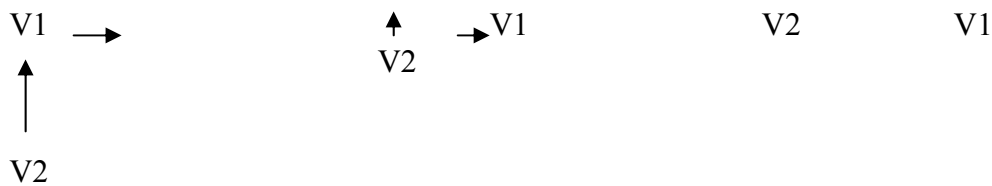


Figure 3.3 – Two potential scenarios of a push-chain relationship between two vowels, V1 and V2, where V2 initiates the chain shift.

Under scenario 1, the initiating vowel V2 raises and thereby encroaches upon the space occupied by V1, which in turn moves out of the way. This essentially implies a pattern of *sequential* vowel movements. Scenario 2, on the other hand, states that raising of V2 implies centralisation of V1 at any given stage of the overall process on some level of manifestation of the vowel shift (i.e. on the level of the individual speaker, or a group of speakers, or all speakers).

Note that since any scientific analysis of a diachronic process can only be carried out on a subsample of speakers, only scenario 1 can be clearly demonstrated. More specifically, the presence of two overlapping vowel distributions refutes scenario 2 for that particular vowel shift, whereas there is more room for speculation if a scenario 2/stage 2 relation is observed (i.e. this may have been preceded by an overlapping phase itself, for which in that particular case we simply do not have evidence).

All of the above points are essentially open to empirical investigation, and the present chapter will address each of them in turn. One further topic around which there has been some debate relates to the spatio-temporal relations of the NZE SFV shift. More specifically, it was not clear whether all three steps of the shift are endemic and have occurred after the

arrival of the first settlers in New Zealand or whether raising and fronting of DRESS/TRAP and centralisation of KIT was already present in the speech of those speakers. Earlier analyses have taken polar views, which means that they have adopted a viewpoint of either exclusive innovation in NZE (which is the view of Bauer 1979, 1992) or exclusive conservatism vis-à-vis British English (Trudgill 1986). Later analyses have arrived on the more reconciliatory conclusion that raised/fronted variants of DRESS and TRAP were already present in the speech of the earliest settlers, and that this process continued in New Zealand (as well as in the other Southern Hemisphere varieties) afterwards (Trudgill, Gordon and Lewis 1998). In addition, KIT centralisation has been analysed as an endemic phenomenon (Gordon et al. 2004). There are relatively few examples of centralised KIT in Gordon et al.'s (2004) study of the first and second generation of NZE speakers (1850 – 1900). Thus, it seems that the crucial period for KIT centralisation was the subsequent Intermediate Period, which will be discussed here.

Although the body of studies on NZE SFVs is now considerably large, almost all of them focus on the *qualitative* properties of the vowels involved in the shift. This may, however, be only a part of a more inclusive process in which the duration dimension (i.e. vowel *quantity*) plays a subsidiary role. The main reason for this seems to be that it has been assumed *a priori* that in terms of length/duration all three SFVs are in some meaningful sense ‘the same’ (hence the designation as ‘short front vowels’), and that the chain-shift involves only a changeover in the spatial arrangements of KIT, DRESS and TRAP<sup>20</sup>. In section 8 of this chapter, I will also analyse the durational properties of these vowels and argue that the duration dimension may illuminate a number of observed properties in the F1/F2 dimension. Furthermore, I will argue in chapter 8 that this assumption has broader implications for the study of vowel movements, both methodological as well as theoretical.

Finally, it should also be noted that the NZE short front vowel shift stands out as a true exception to the general principles of chain-shifting outlined by Labov (1994), who concedes that ‘this [the NZE short front vowel shift] clearly violates Principle III (since a short front

---

<sup>20</sup> This relates both to the studies on NZE front vowels (for exceptions to this rule see Mackenzie 2005/Maclagan and Hay 2005) as well as to sociophonetic studies on vowel movements in general, where the qualitative dimension has received overwhelming attention. The reason for this seems to be that the qualitative dimension lends itself more readily to auditory analysis (as well as to historical reconstruction based on synchronic phonemic correspondence), which was the approach to analysing vowel change until recently. Although acoustic analysis may have changed the methodology of empirical investigations on vowel systems, it has in many cases continued investigating the basic issues as laid out by previous approaches, i.e. analyses of vowel quality.

vowel is moving to the back in a chain shift) and Principle II (since short front vowels are rising together) ‘ (p 138). This issue will be discussed at length in chapter 7.

### 3.3 Results - overall patterns

As figures 3.4 (a) - (b) show, the global pattern of change within the intermediate sample is fairly consistent. The younger speakers have higher realisations of DRESS and TRAP as well as a more central realisation of KIT, which matches the overall developments in the SFV system of NZE over the last 150 years (For comparison, figures 3.4 (c) – (d) show SFV averages for the Mobile Unit (MU) speakers analysed in Gordon et al. (2004). Figures 3.4 (e) – (f) shows modern speakers (cf. Maclagan 1982). The data on which figs. 3.4 (c) – (f) are based were kindly made available to me by Margaret Maclagan). The only discontinuity is in the behaviour of the KIT vowel in the speech of the LATE MALES, who have a less central realisation than the corresponding vowel in the medium male sample. For the males, the decrease in the F2 dimension for the KIT mean within the Intermediate sample is in the order of 256 hz. Given the simultaneous raising/fronting of DRESS (-52 hz in the F1 and 153 hz in the F2 dimension), this brings about a change that increases both the distance between KIT and DRESS<sup>21</sup> as well as shifting the basic typology of the two vowels relative to each other. That is, whereas the difference between KIT and DRESS was one of height for the earlier speakers, especially the males, it develops into a system where both vowels are of approximately the same height, but contrast on a front-back (or, in acoustic terms, F2) dimension.

---

<sup>21</sup> Henceforth, the term ‘distance’ will be used in an empirical (i.e. mathematical) sense as an expression of Euclidean distance between two points in linear two-dimensional space, such as a vowel plot in hertz.

### Intermediate Male Speakers - SFV averages

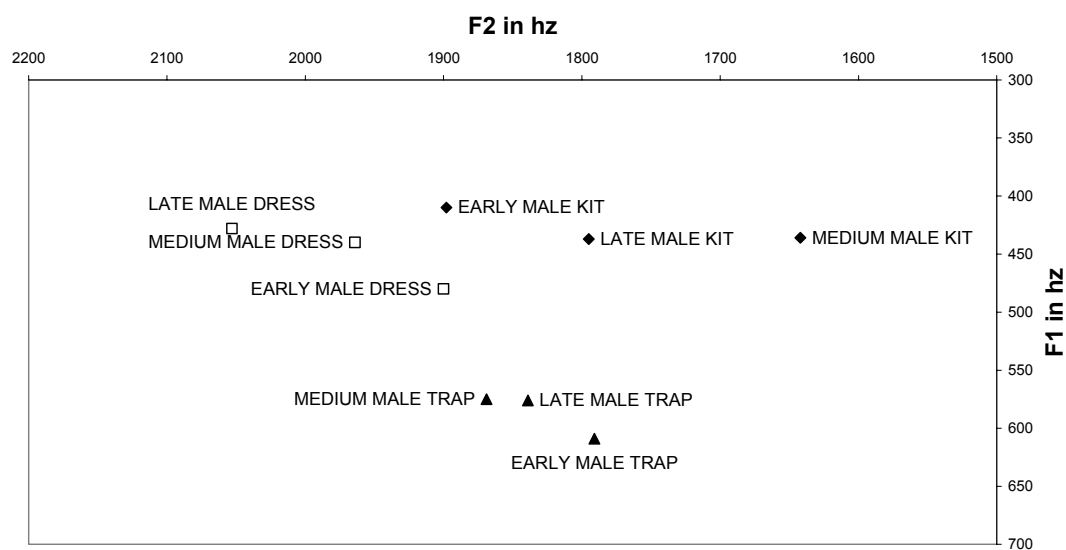


figure 3.4 (a)

### Intermediate Female Speakers - SFV averages

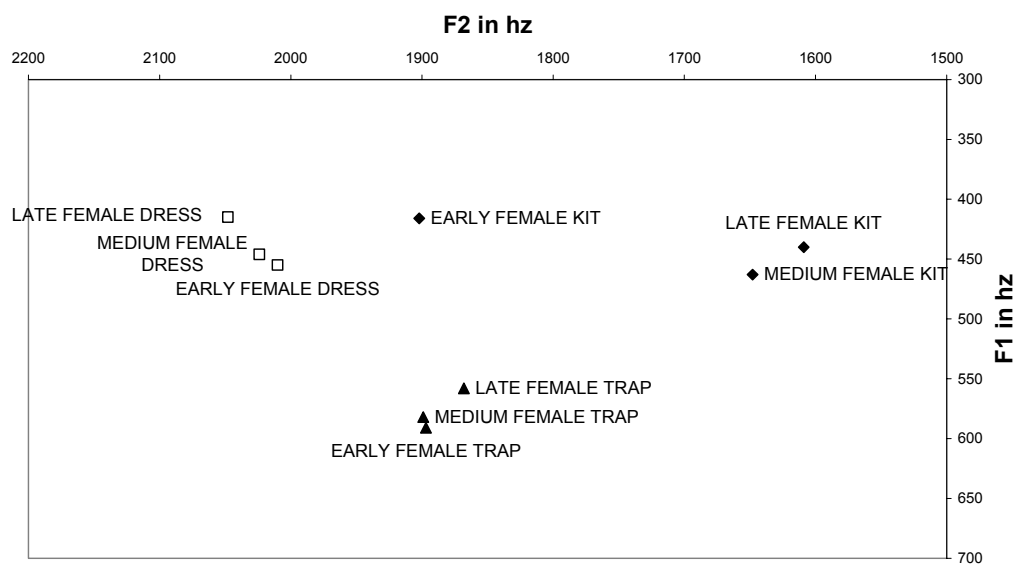


figure 3.4 (b)

### Mobile Unit Males (pre-Intermediate)

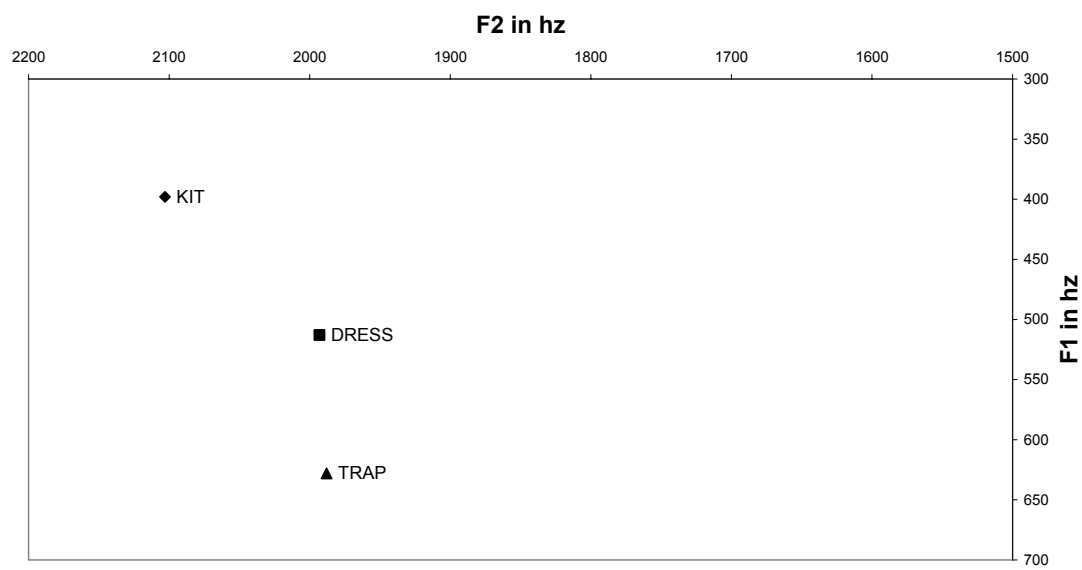


figure 3.4 (c)

### Mobile Unit Females (pre-Intermediate)

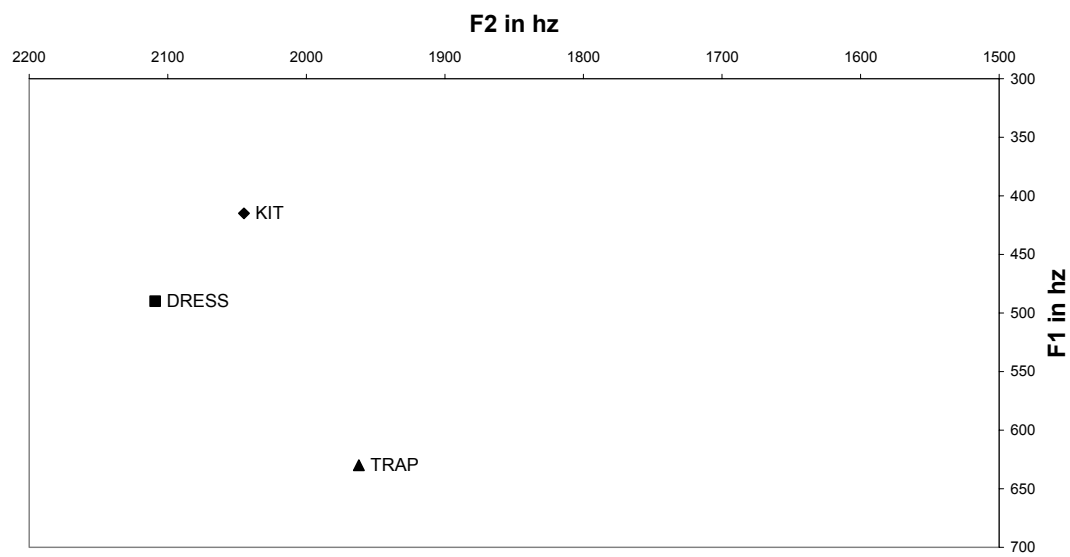


figure 3.4 (d)

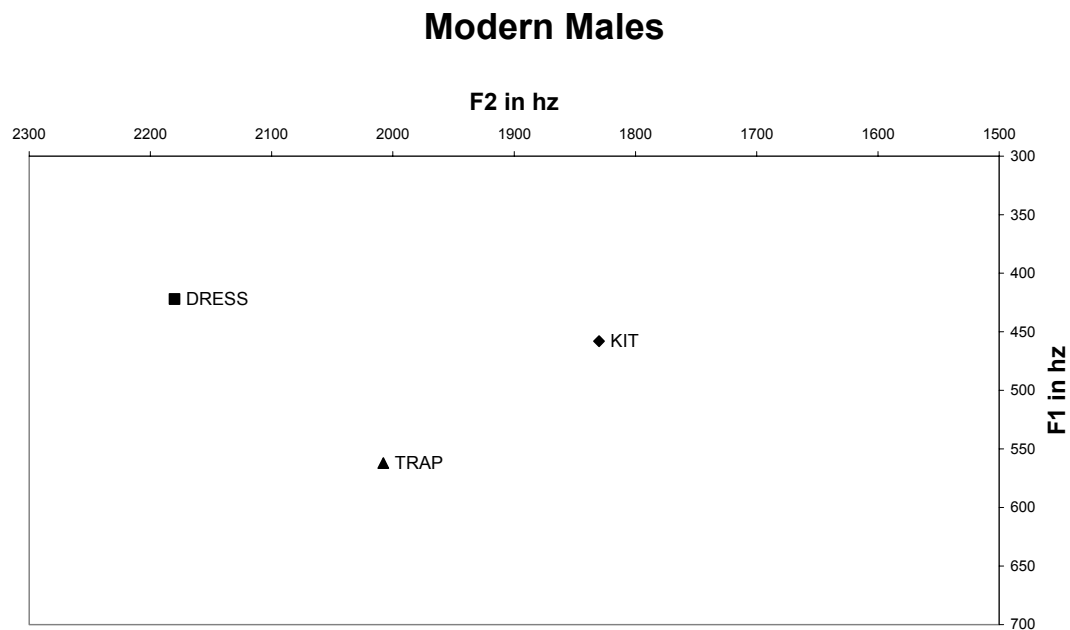


figure 3.4 (e)

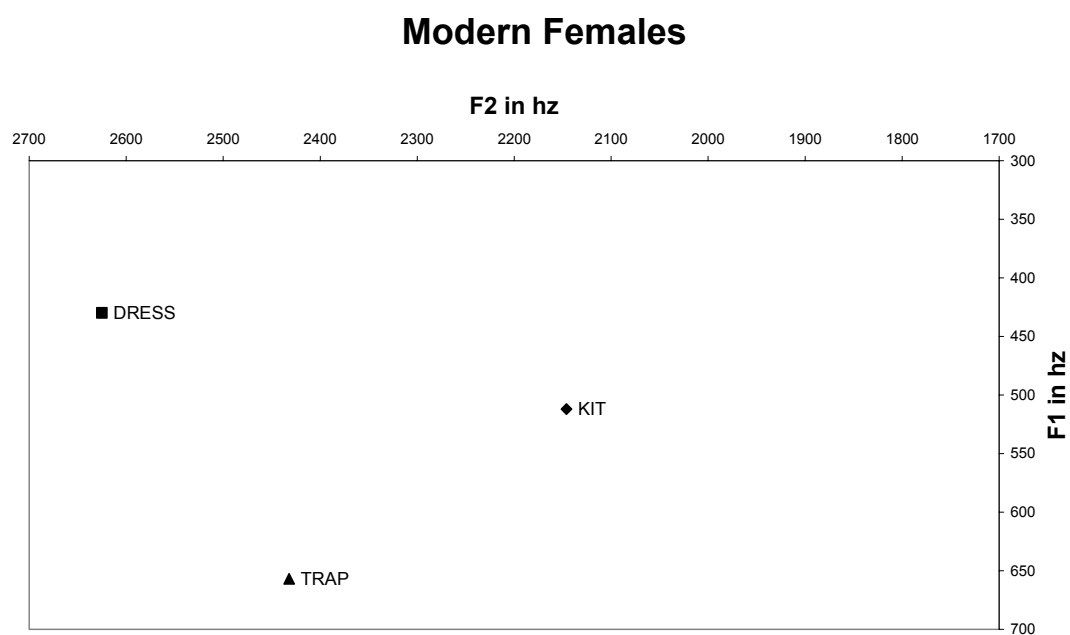


figure 3.4 (f)

Fig 3.4 (a) – (f) – F1/F2 averages of KIT, DRESS and TRAP for all groups of Intermediate speakers (figs. 3.4 (a) – (b), pre-Intermediate speakers (3.4 (c) – (d)) and modern speakers (3.4 (e) – (f)) (cf. Gordon et al. (2004) and Maclagan (1982). Note that the data plotted in figs 3.4 (e) and 3.4 (f) are non-normalised. High F1 values correspond to openness. Similarly, the F2 axis corresponds to the front-back dimension, where higher F2 values correspond to more fronted articulation.



## Early Males - SFV Averages and contextual variants

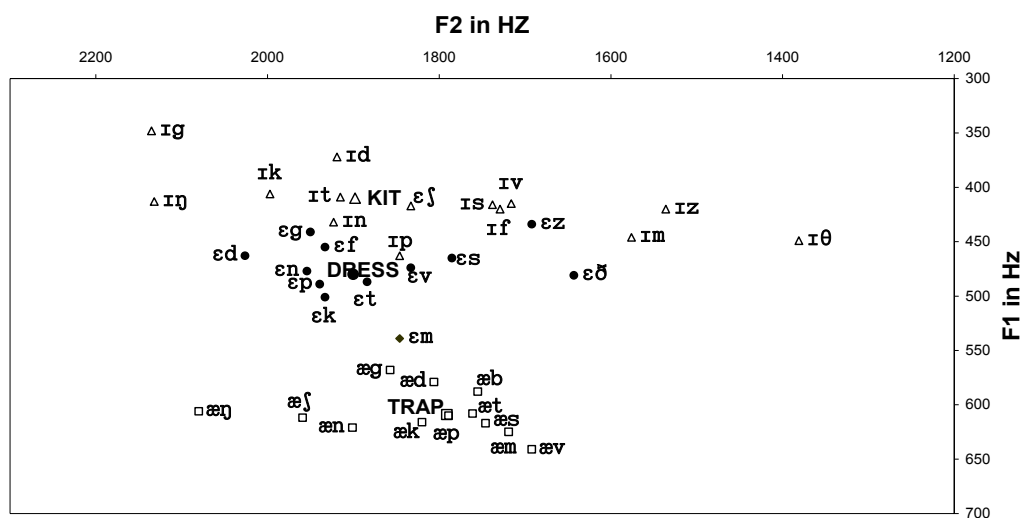


figure 3.5 (a)

## Medium Males - SFV Averages and contextual variants

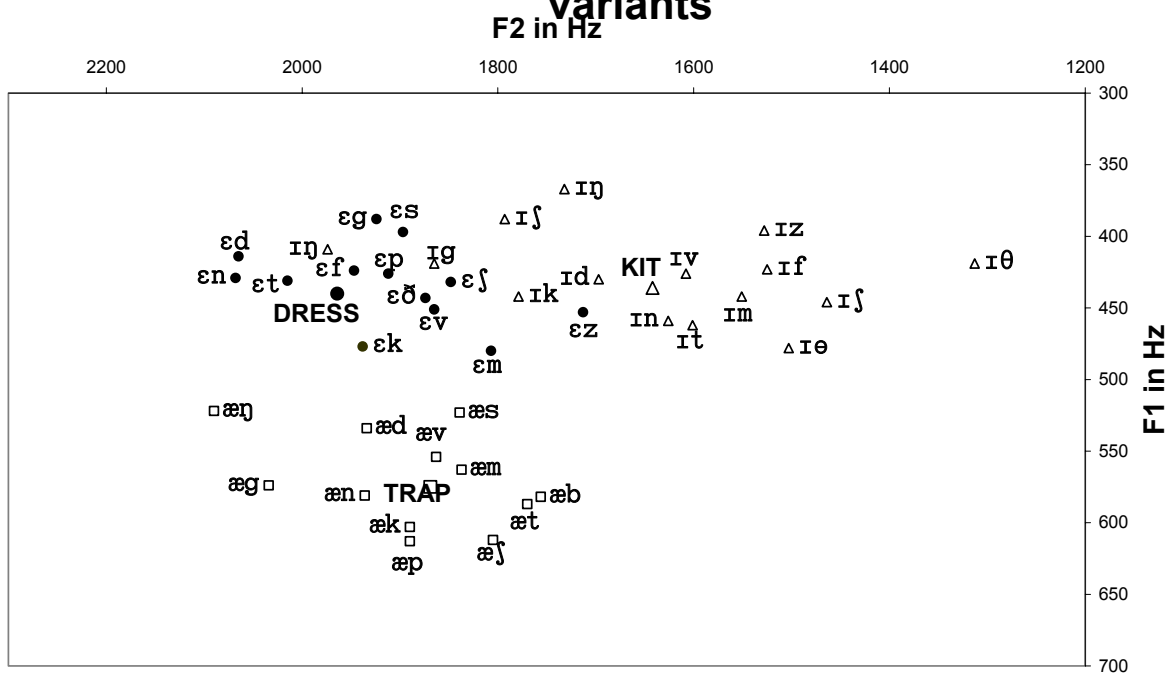


figure 3.5 (b)

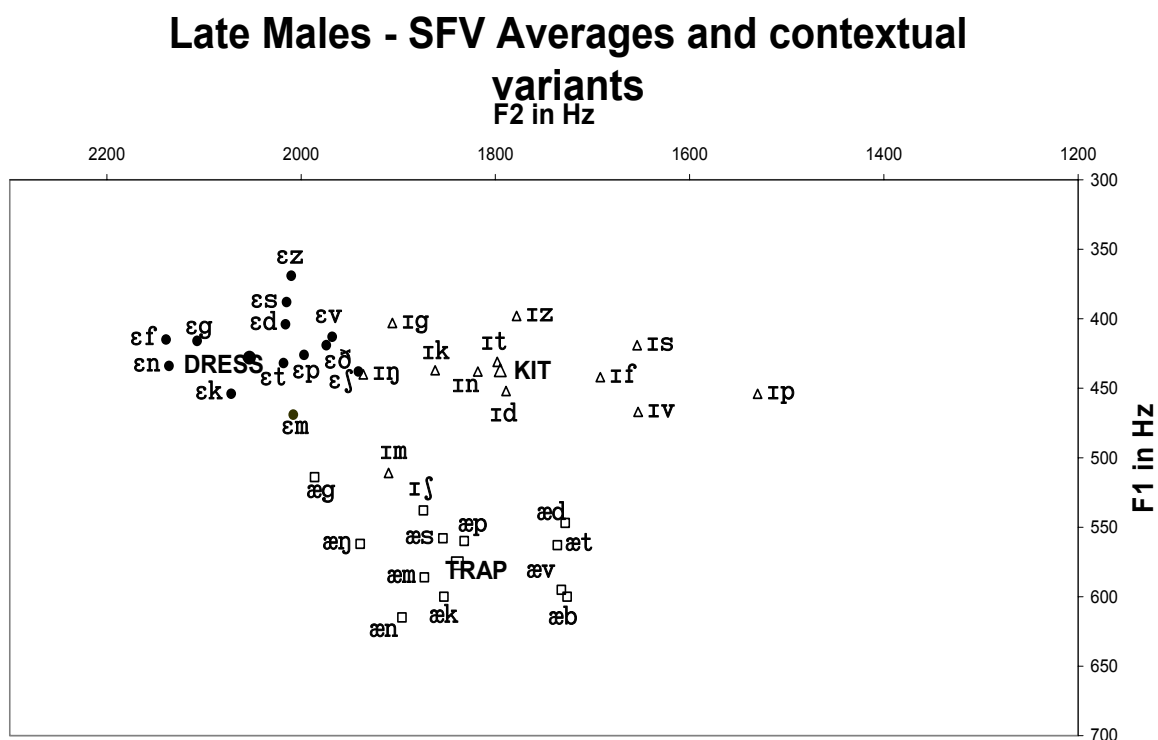
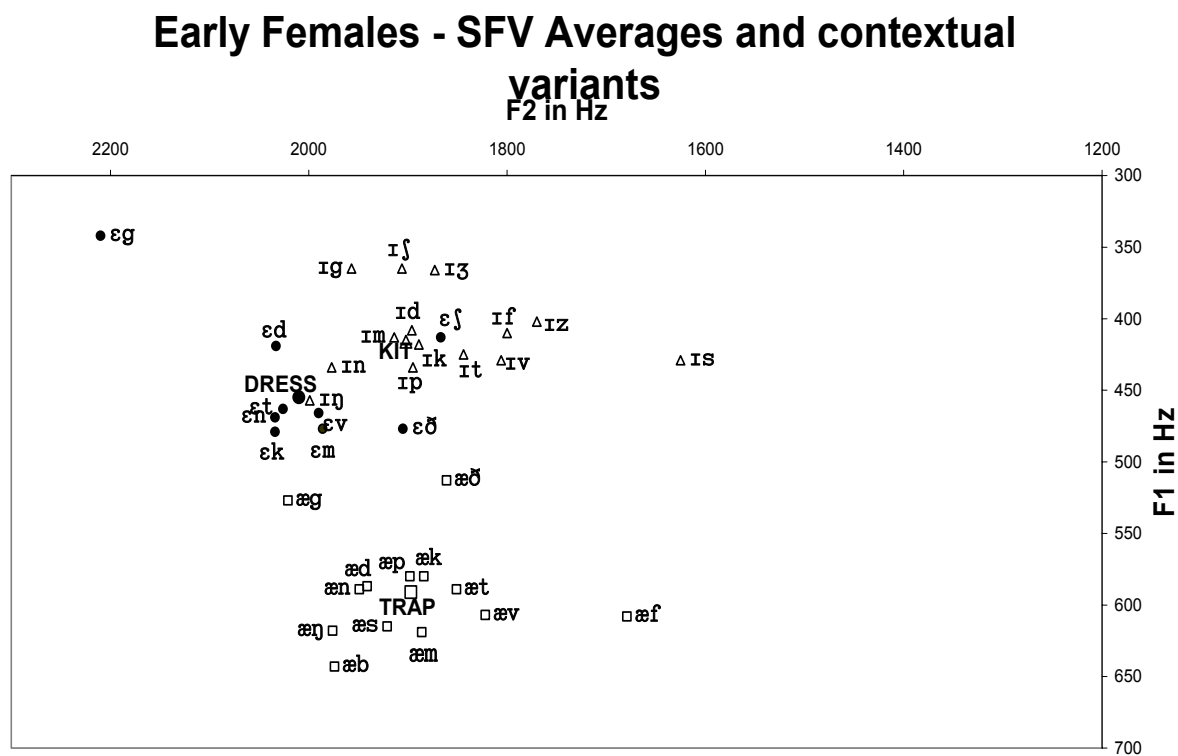


figure 3.5 (c)





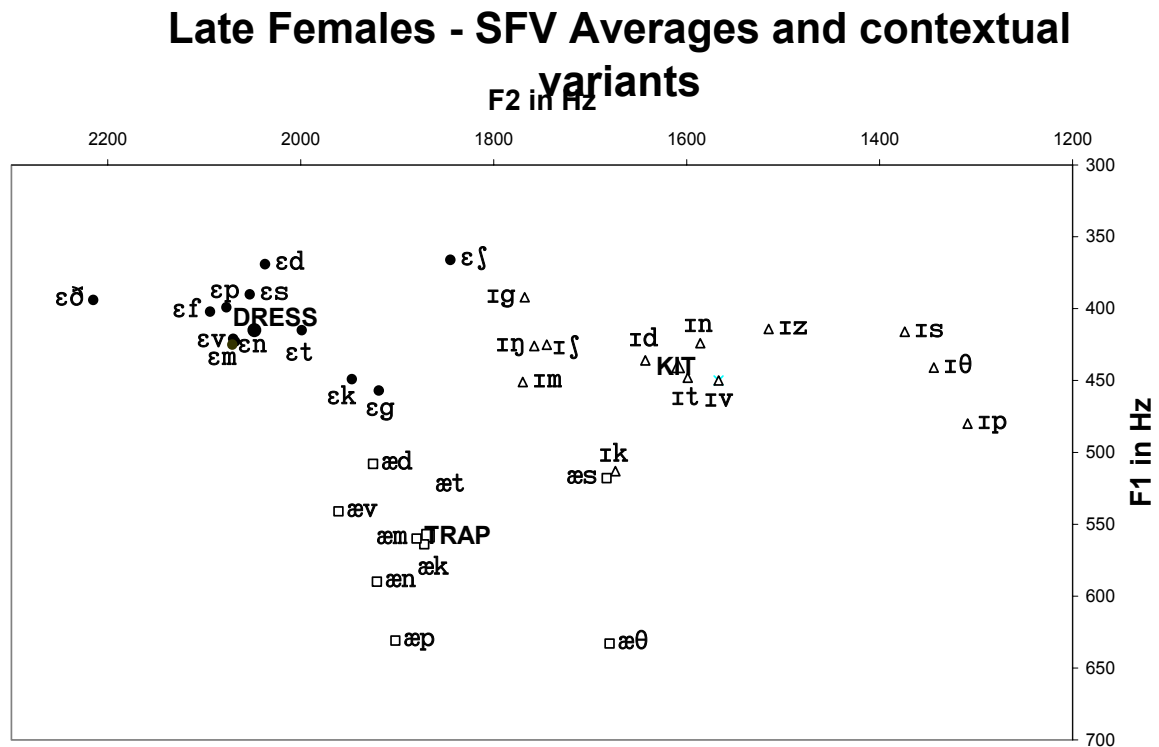


figure 3.5 (f)

Figure 3.5 (a) – (f) – Mean frequency values of the first two formants of the lexical sets of KIT, DRESS and TRAP as well as their contextual variants. Each of the six groups of Intermediate speakers is plotted.

Figure 3.5 plots the position of the three SFVs and their respective contextual variants. Only those categories that occurred at least five times within any subgroup are plotted. Impressionistically, three characteristics stand out: First, there is substantial categorial overlap between a number of contextual variants in the lexical sets of KIT and DRESS in the speech of the EARLY and MEDIUM MALES as well as the EARLY FEMALES. Second, there is an obvious stretch in the categories of KIT in the F2 dimension in the speech of the EARLY MALES. Third, there is no categorial overlap between the lexical sets of DRESS and TRAP for any group of speakers in the sample.

### 3.3.1 Evidence for a push chain scenario

Table 3.1 shows the Euclidean distances between the mean values of the SFVs for all six groups of speakers.

	KIT <-> DRESS	DRESS <-> TRAP
EARLY MALE	70.02	168.88
MEDIUM MALE	322.02	165.07
LATE MALE	258.16	260.19
EARLY FEMALE	114.83	176.82
MEDIUM FEMALE	376.38	184.71
LATE FEMALE	439.71	229.89

Table 3.1 – Euclidean distances between the mean values of the lexical sets of KIT, DRESS and TRAP in the speech of the six groups of Intermediate speakers.

The numbers indicated in table 3.1 suggest a fairly clear picture. The main jump in terms of Euclidean distance between KIT and DRESS occurs between the early and the medium age groups for both genders, whereas DRESS and TRAP move apart from each other only later, i.e. between the medium and the late stage. This behaviour would suggest a pull-chain scenario, if it weren't for the following points: In the speech of the early speakers, the distance between KIT and DRESS is extremely small, and there is substantial overlap of various allophones of both lexical sets (cf. fig. 3.5). Thus, it is hard to see how such a system could be viewed as an initial state in the vowel shift. In addition, a pull-chain scenario would probably assume that the second step in the shift was the movement of DRESS to a higher position, followed by TRAP. However, this seems unlikely, given the consistent increase of the distance between DRESS and TRAP over time in the sample (under the pull chain assumption, you would probably expect an initial increase, followed by a decrease of distance between the two lexical sets). We would also expect to find speakers with centralised KIT, but not-yet raised DRESS under this scenario. As will be shown in section 3.7, this is not the case, whereas the reverse possibility holds. It is therefore more likely that the system of the early speakers is a transitional one where DRESS has already been raised sufficiently to trigger a reactive centralisation of KIT. With respect to DRESS and TRAP, it seems reasonable to extrapolate an earlier stage where the two lexical sets were closer to each other.

### 3.3.2 The ‘Split System’ in the EARLY MALE sample

It is obvious from looking at the above plots that the early male group of Intermediate speakers stands out with respect to both the overall means of DRESS and KIT as well as the structural set-up of the contextual variants. They are the only group that has allophones of KIT as the most fronted tokens, but they also have centralised variants, which leads to a striking ‘stretch’ in the F2 dimension for KIT. It seems clear that the KIT items that occupy the front part of the overall distribution are those that are followed by a velar consonant. What is less clear, however, is whether the front categories represent variants that have not yet undergone centralisation, or whether this early group of speakers shows a strategy of both fronting and raising pre-velar variants of KIT and simultaneously centralising others<sup>22</sup>. My auditory assessment of those items suggested that there are in fact a number of tokens that are similar to the realisation of KIT in AusE (i.e. as [i]), as well as some that are closer to the RP value ([ɪ]). This impression is supported by the data given in Bernard (1970), which states mean formant frequency values of 365 hz for F1 and 2220 hz for F2 for modern AusE KIT, which is in the vicinity of the EARLY MALE realisation of pre-velar KIT. Secondly, a similar system is described by Lass (1987) and Wells (1982) for modern South African English, although the conditioning factors differ (Lass reports as the main conditioning factors for KIT fronting/raising: KIT in initial position, after /h/, and next to velars/palato-alveolars).

One obvious point that needs to be made clear in this context is that the distributional properties (i.e. stretch and overlap) of the SFVs as described above are not merely an artefact of pooled data from various speakers of differential degrees of innovativeness. Figure 6 below plots all measured KIT/ DRESS tokens of J.M., an early male speaker.

---

<sup>22</sup> It should also be noted that there is a striking mismatch in terms of conditioning factors in the arrangement of the contextual variants of KIT in that group. Whereas the front categories share the same place of articulation (velar), the centralised ones are predominantly followed by fricatives (i.e. manner of articulation). The implications of this observation will be discussed below.

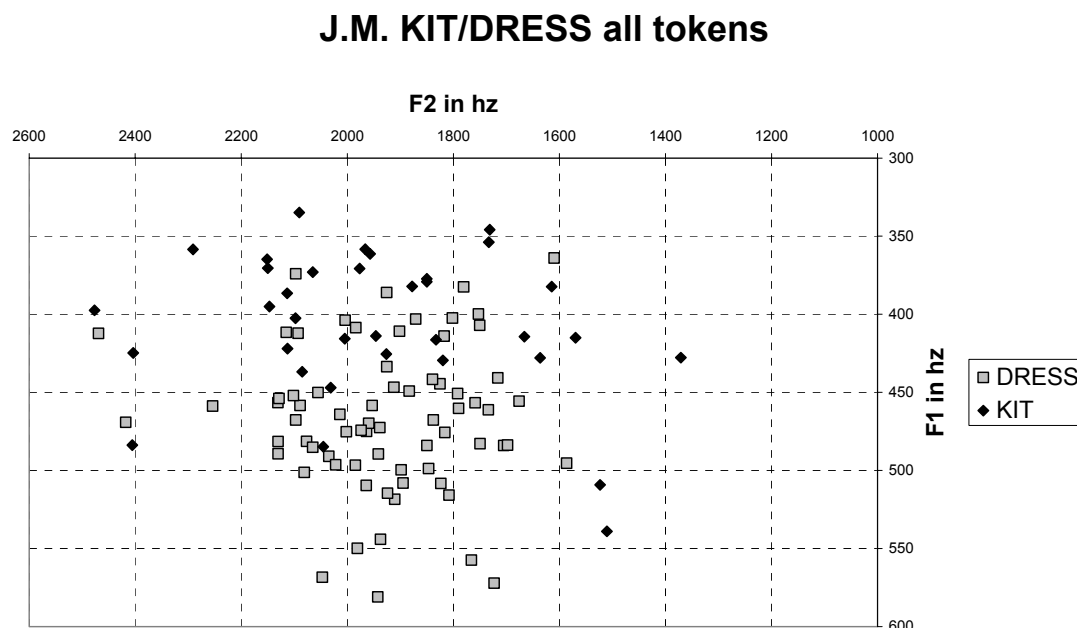


Figure 3.6 – F1/F2 overlap in the speech of an individual speaker from the early male sample.

Figure 3.6 shows that (a) there is considerable overlap between DRESS and KIT in F1/F2 space even in the speech of individual speakers, and that (b) the clustering of DRESS tokens around their F2 mean is somewhat tighter than is the case in KIT (the standard deviation in F2 for that speaker is 263.9 hz around an average of 1943.7 hz in KIT, and 167.2 hz around 1930 hz in DRESS).

However, it is less clear how an elliptical system such as that of the EARLY MALES should shift at all, rather than merge. It has been recognised that a given vowel (i.e. as an etymological category) does not necessarily behave as a coherent category in a vowel shift (cf. M. Gordon 2001, 2002 for a summary as well as Labov 1994/2001 for ample evidence of transitional allophonisation of vowels which undergo a shift). Rather, different contextual variants can shift at different rates, or even in different directions. This seems to have happened at the early stages of KIT centralisation in NZE, however, the modern resolved system has restored a uniform etymological category of KIT.

### 3.3.3 Discussion – Means and allophones

We have seen that it was in the Intermediate Period where the shift in the short front vowels came to its completion, resulting in the modern set-up with two front vowels and a central

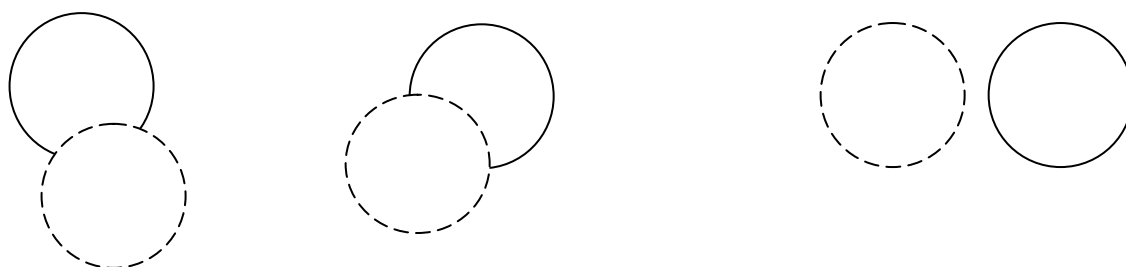
vowel. In addition, it could be demonstrated that there was a stage where the SFV system had the peculiar distributional property of an unusually elliptical distribution in the top two heights (KIT and DRESS), where one segment (KIT) occupies an unusually large space in the front-back (or F2) dimension. This has two implications for the study of vowel shifts, a general one as well as a more specific one (i.e. specific to the study of the NZE short front vowel shift). In general terms, the process of condensing the available vowel space for a given vowel in one articulatory dimension (in our case, the front/back dimension of KIT) in an intermediate stage of a vowel shift imposes limits upon the usefulness of mean values of an entire lexical set or etymological category (such as KIT or DRESS), since these mean values may obscure both the pathways of movement of meaningful smaller units within a lexical set (such as, for example, KIT before alveolar fricatives) as well as certain differences between speaker groups (such as EARLY MALE and EARLY FEMALE, who have similar mean values in the lexical set of KIT, but differ markedly in their degree of consistency around the mean). An analysis that is based solely on overall means cannot capture the difference between the two putative scenarios depicted in figure 3.7 below.

*Stage 1*

*Stage 2*

*Stage 3*

Scenario 1 – Isomorphous distributions throughout the shift





### Scenario 2 – Heteromorphous distributions throughout the shift

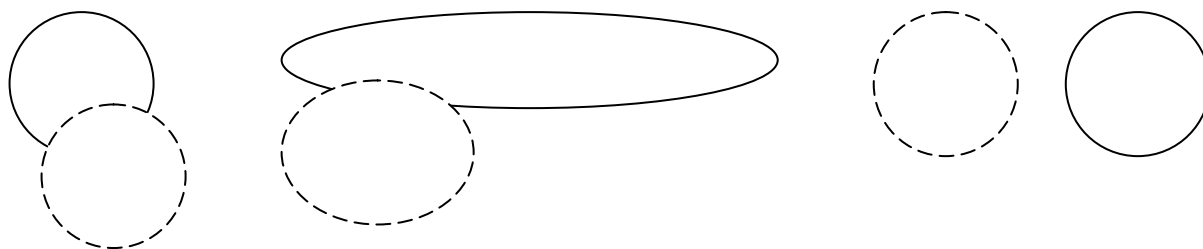


Fig. 3.7 – Two possible scenarios for the NZE front vowel shift in the top two heights. The upper circle (or ellipse) represents an idealised distribution of KIT-vowels (solid line), the lower one that of DRESS (dashed line).

Apart from the general implication of this finding as a demonstration of the existence of highly elliptical vowel distributions in general (as opposed to, for example, an assumption of mutually implied degree of innovativeness within a given sub-group of the speech community that undergoes the shift), this might also link the history of the short front vowel shift in NZE to that of Australian English (AusE), which behaved similarly with respect to DRESS and TRAP, but raised and fronted KIT rather than centralising it. That is, it can be hypothesised that at an earlier stage, both varieties had at their disposal a range of KIT-variants spanning the entire range between (roughly) [i] and [ə], but generalised different means only later (this hypothesis seems to be confirmed by ongoing research on the ‘Intermediate Period’ in AusE. P. Trudgill and E. Gordon, personal communication).

In addition, we have seen that it is through the study of such smaller units that apparent temporal discontinuities can be accounted for as a resolution of formerly overlapping lexical sets rather than the absolute position of a given lexical set (cf. KIT in the MEDIUM and LATE MALE groups).

Another point that I would like to raise in relation to the short front vowel shift in NZE is of a more terminological nature and relates to the notions of ‘KIT centralisation’ or ‘centralised KIT’. Heretofore, this term has been used in a rather loose fashion in order to explain both the process as well as the outcome of a historical process. Given the nature of the short front vowel system of NZE in the early Intermediate Period, those terms acquire a certain amount of denotational ambiguity, in that it is not clear whether what is referred to in any given instance of usage of these terms is a centralised KIT mean of the overall lexical set

or the existence of centralised variants in the speech of any subgroup in the speech community (an age or gender cohort, a social class, an individual). That is, the KIT vowel in the speech of a given sub-group or an individual can be both non-centralised (in that it has a mean position right ‘on top’ of DRESS) and centralised (with the simultaneous existence of central allophones) at the same time, which implies that some conceptual clarification is probably advisable in future studies on this topic.

On a more sociolinguistic note, two things have to be pointed out regarding the role of the female speakers in the sample. First of all, they tend to clearly be spearheading the chain-shift in most of its various dimensions. That is, any given female group has more progressive mean values (more centralised for KIT, more raised for DRESS/TRAP) than their corresponding male counterparts of the same age cohort (cf. Figure 3.1-5). This confirms the tenet of the leading role of women in non-stigmatised sound change, i.e. sound change ‘from below’ (Labov 2001). On the other hand, the above analysis should have made clear the limited usefulness of overall mean values of lexical categories in an instable system. For example, both the MALE and the FEMALE speakers of the EARLY group have rather similar mean values for KIT, but differ considerably in the arrangement of the contextual categories (cf. Figure 3.5).

Whereas the EARLY MALES have a system where KIT can be either the most fronted element as well as the most central one of the SFVs, this does not hold true for the EARLY FEMALES, who are much more consistent and do not show the same degree of allophonisation. For these two groups of speakers then, it is the MALES who have both the most conservative as well as the most innovative realisations of KIT, which brings about a cancelling out around a mean that is close to that of the females.

### 3.3.4 Factor analysis

#### *Coarticulation*

Apart from pointing out the distributional properties of the SFV during the shift, we have mentioned in passing that these distributions do not represent random clusterings of KIT/DRESS/TRAP tokens, but tend to have internal structure depending on the phonemic environment they occur in. I will briefly review the effects of adjacent phonemes on vowels, before moving on to whether these expected patterns show up in the sample. With respect to place of articulation, Stevens and House (1963) have found that:

In the environment of front vowels, for example, ‘velar’ consonants (being palatal variants in English) have a high F2-locus (above 200 cps) whereas the F2-loci for postdental and labial consonants are below the F2 values for the vowels (p. 125).

We would therefore expect a distribution where pre- and post-velars occur towards the ‘front end’ of a distribution, whereas vowels in the other environments are more central. With regard to F1 effects, they found both a displacement of vowel frequencies towards those of the adjacent consonants (which is always a downward shift) and a shift towards a neutral value (i.e. that of schwa, in the vicinity of 500 hz) if the articulatory target of the vowel is far from that of the surrounding consonants.

As for manner of articulation, they state that:

One feature of the data [...] is the tendency for F2-values for vowels in the environments of fricative consonants to be lower for front vowels and higher for back vowels relative to corresponding values for stop consonantal environments. This difference is most evident in the vowels /ɪ/, /ɛ/, /æ/ and /ɑ/ (p. 126).

In addition, Wright (1986) has found a shrinking of the overall perceptual vowel space for vowels in nasal environments.

For a more comprehensive survey of the relationship of vowel targets and phonemic context, see chapter 4 of Harrington and Cassidy (1999).

*CART (classification and regression tree)*

As a second step in the analysis, all tokens were coded for the independent variables outlined in table 3.2 below, and subsequently subjected to a CART (classification and regression tree) analysis.

<i>CODING CATEGORY</i>	<i>VALUES</i>
SPEAKER	Each individual from the sample of 30 speakers
AGE	EARLY/MEDIUM/LATE
GENDER	MALE/FEMALE
ETYMOLOGICAL CATEGORY	KIT/DRESS/TRAP
FOLLOWING CONSONANT	The following phoneme
PLACE	Place of articulation of the following consonant – LABIAL/DENTAL/ALVEOLAR/POST-ALVEOLAR/VELAR
MANNER	Manner of articulation of the following consonant – STOP/FRICATIVE/NASAL/LIQUID
VOICING	Voicing feature of the following consonant – VOICED/UNVOICED
SYLLABICITY	The syllable structure of the word in which the token occurs – COMP(i.e. stressed syllable in a compound word)/σ/σσ/σσσ/σσσσ/σσσσσ/σσσσσσ <sup>23</sup>
PRECEDING CONSONANT	The preceding phoneme
PRE-PLACE	Place of articulation of the preceding consonant – LABIAL/DENTAL/ALVEOLAR/POST-ALVEOLAR/VELAR/GLOTTAL
PRE-MANNER	Manner of articulation of the preceding consonant – STOP/FRICATIVE/NASAL/LIQUID/GLIDE
PRE-VOICING	Voicing feature of the preceding consonant – VOICED/UNVOICED
minimal1	whether there is an <i>exact phonological</i> minimal pair in one of the other SFVs.
minimal2	whether the lemma has a minimal pair in one of the other SFVs.
minimal3	whether there is a minimal pair in one of the other SFVs in the <i>same part of speech</i> .

Table 3.2 – Coding categories. Each token has been coded for the above categories and subsequently analysed using the CART component of the R statistics program.

<sup>23</sup> The bold sigma here stands for the syllable containing the measured token.

*Phonetic factors – F1*

Figure 3.8 (a) below shows a CART representation of pooled F1 measurements regressed onto the following independent variables: *age*, *sex*, *vowel*, *F2*, *pphon*, *pplace*, *pmanner*, *pvoice*, *phon*, *manner*, *place*, *voice*, *syll*, *minimal1*, *minimal2*, *minimal3*. The tree is truncated below the third branching point. 3.8 (b) – (c) shows the data for each of the three vowels regressed onto the same variables. The following conventions are used: At the top of each branching point, the independent variable that splits up the data at that branching point is given in bold. At the bottom of each node, the average value of each respective subset of the data is indicated in italics. The values of each of the subgroups within each independent variable are indicated next to the branch.

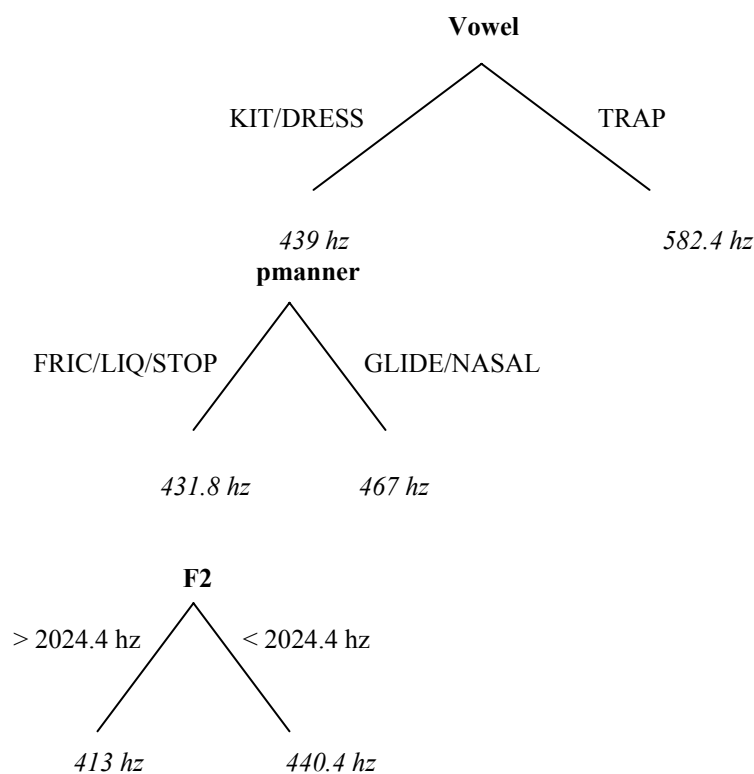


Figure 3.8 (a) – CART analysis of pooled F1 measurements of all three SFVs.

Not surprisingly, the major distinction in the F1 data is between the TRAP data on the one hand, and the KIT/DRESS data on the other hand. Furthermore, the latter branch is divided up

into a range of allophones that depend on the manner of articulation of the preceding consonant, where vowels following nasals and glides show a markedly lower F1 than do vowels in other environments. Within that latter group, we note a subdivision whereby the lowest F1 values are found in the data pool that has the higher F2. This point, in conjunction with the fact that age does not appear in the tree on this global scale, implies that (a) the interrelation of raising and fronting in DRESS as well as lowering and centralisation in KIT is found only in some allophones and that (b) this seems to be rather a property of individual systems, which, however, brings about an overall correlation in the entire sample between F1 and F2 in DRESS (cf. section 3.7 below).

Within the KIT data (figure 3.8 (b)), the major predictor of F1 is F2, where the same relationship holds: vowels with higher F1s show lower F2s, and vice versa. The interpretation for this would be that fronted (or at least non-centralised, cf. above) allophones of KIT are expected to be high in order to stay clear of the space occupied by DRESS, whereas this is not the case for the more central tokens.

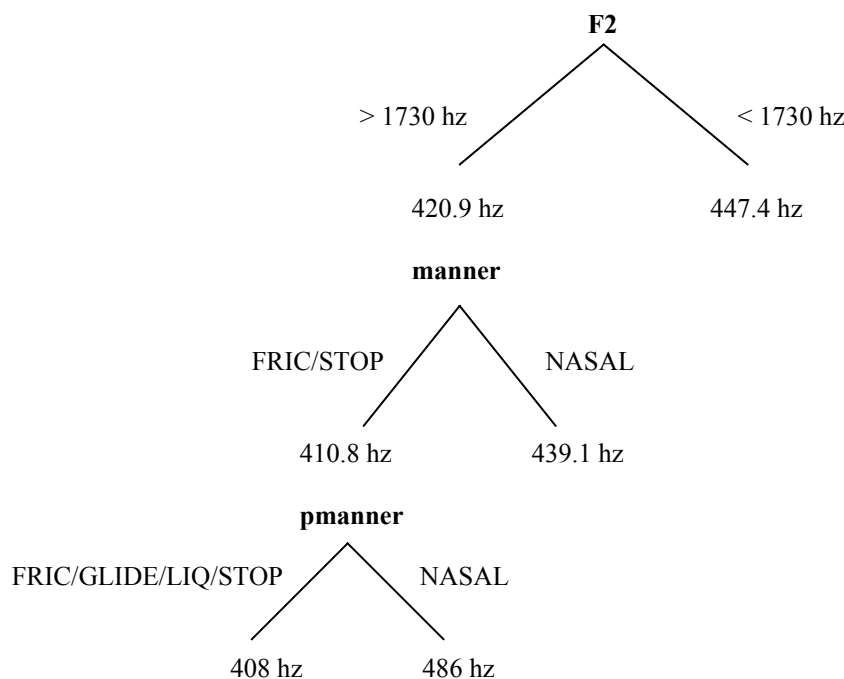


Figure 3.8 (b) – CART analysis of all KIT F1 measurements.

The condition whereby vowels in nasal environments show higher F1 values holds in the KIT data as well. Interestingly, however, this holds only within the branch of higher F2 values,

from which we may question the above finding that if anything, markedly lower allophones of KIT would be found predominantly for the central branch. However, upon inspection of the DRESS data (figure 3.8 (c)) it seems clear that the comparable allophones in the latter vowel also show comparatively high F1 values and that, in addition, height is a function of F2 whereby the highest post-nasal DRESS vowels are found within the branch that has the lower overall F2, i.e. the branch which is closer to KIT in the front/back dimension. This seems to suggest an interplay between coarticulatory constraints as well as systemic requirements: Post-nasal allophones of DRESS show markedly higher F1 values than their non-nasal counterparts (probably due to coarticulation) as well as being sensitive to their counterparts within an adjacent distribution.

Unlike KIT, however, age does show up in the DRESS data, if only at the terminal nodes and only within the class of non-nasal allophones. Overall, we find the expected relationships: Late and medium speakers have lower F1 values than early speakers. More interestingly, this particular subgrouping recurs throughout the entire sample and suggests a somewhat leap-like pattern in the restructuring of the SFV system, whereby the early subsample is closer to the initial state, whereas the medium and the late speakers are more similar to the current state (cf. fig. 3.1 above).

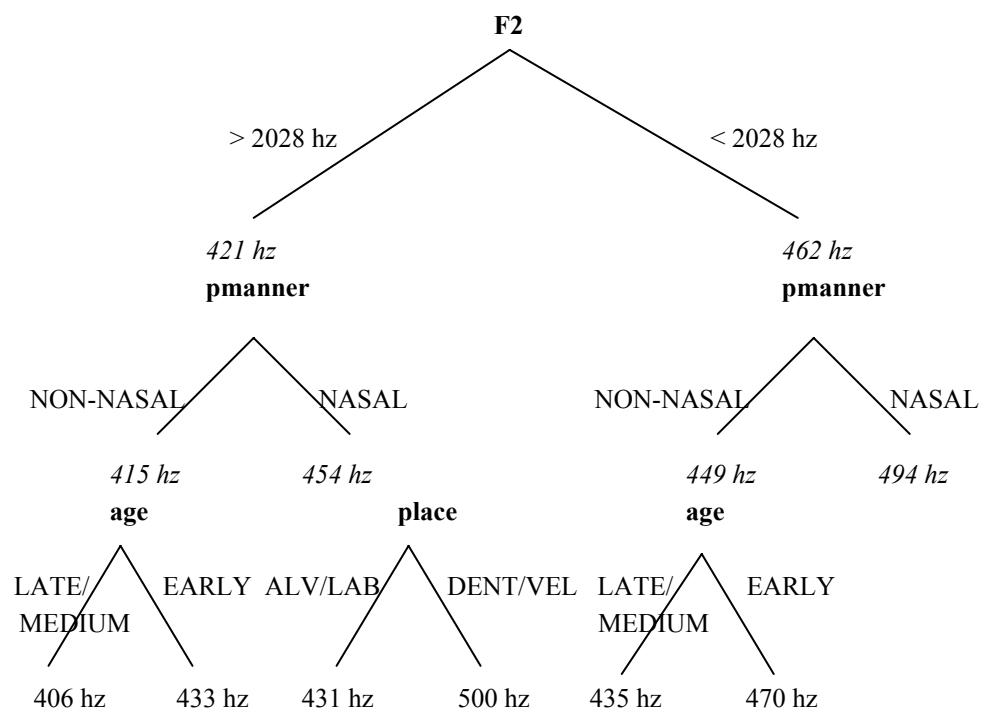


figure 3.8 (c) – CART analysis of all DRESS F1 measurements.

In contrast to KIT and DRESS, the TRAP data (figure 3.8 (d)) shows no stratification in terms of phonemic environment or F2. Rather, the first-tier branching splits up the data into age groups, with early speakers showing higher F1 values than late and medium speakers, which reflects the subgrouping in innovativeness we observed above. Within the late/medium branch, there is a minimal pair effect whereby TRAP tokens that have a minimal pair with *both* KIT and DRESS have lower F1s than do those with a minimal pair in either one of the other two vowels, or none at all.

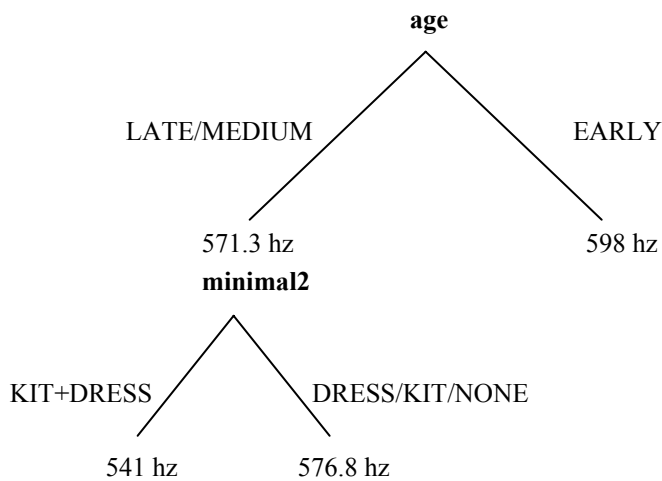


Figure 3.8 (d) – CART analysis of all TRAP F1 measurements.

## F2

In this section we will analyse the F2 data regressed onto the same variables as above (with the obvious difference that F1 is now an independent). Figure 3.9 (a) shows the overall patterns: The major branching splits off DRESS (with high F2s) from KIT and TRAP, which suggests that on the whole, the Intermediate period is indeed intermediate with respect to the advancement of the SFV shift (if we think in terms of the idealised representations as outlined in figure 3.1 above, we would expect different subgroupings at either stage, namely



no subgrouping at the initial stage<sup>24</sup> compared to DRESS/KIT vs. TRAP at the final stage). Here we find a situation where KIT is near the F2 space of TRAP while undergoing centralisation, whereas DRESS has already undergone some fronting, i.e. movement away from TRAP.

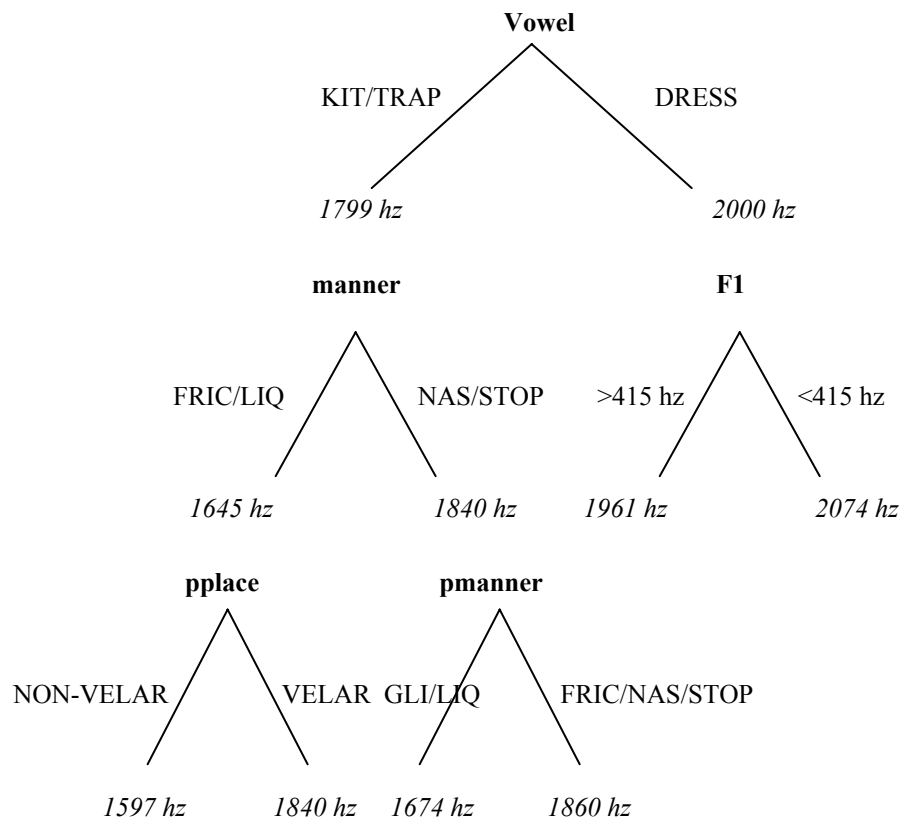


Figure 3.9 (a) – CART analysis of pooled F2 measurements.

<sup>24</sup> This of course depends on how likely a purely structural (i.e. purely vertical) representation as in fig. 1 is assumed to hold in actual data, where by and large we would expect to find at least some positive correlation between F1 and F2 in front vowels (cf. the large-scale plots in e.g. Peterson and Barney 1952). However, even in this case we would not expect the pattern observed in 9 (a) above, but rather a subgrouping of either KIT/DRESS vs. TRAP or DRESS/ TRAP vs. KIT.

Within the DRESS branch, we again observe a pattern whereby tokens with low F2 show higher average F1 values (more on this below). A somewhat complex pattern of allophonisation holds within the KIT/TRAP branch: Vowels before fricatives and liquids (this effect is largely carried by pre-fricative environments since the overall number of pre-liquid tokens is small) show markedly lower F2 values than tokens before nasals and stops. The former group is further split up into a group of post-velar tokens with higher F2 values than their complementary allophones (which is an expected pattern on coarticulatory grounds, cf. section 3.5.4 above). Tokens before nasals and stops are split up depending on manner of articulation of the preceding consonant, where vowels following glides and liquids show lower F2 values than vowels following nasals, fricatives and stops.

Turning to the KIT data (figure 3.9 (b)) we observe an age-group difference with early speakers showing higher F2 values than do late and medium ones, which suggests an appreciable degree of KIT centralisation over the intermediate period. The sub-ordinate branchings show allophonisation along phonemic variables. More specifically, vowels before fricatives show lower F2 values than vowels before nasals and stops in both groups. Again, this is expected on coarticulatory grounds, as e.g. Stevens and House (1963) have explained lower F2 values of vowels before fricatives as a function of undershoot in front vowel articulations. Here, however, the picture is somewhat complicated by the second-tier branching within the late/medium group, where tokens preceded by fricatives and stops show higher F2 values than those preceded by glides, liquids and nasals. With respect to the relationship between nasals and fricatives, we therefore observe a relationship that is the inverse of what was found in the following environment.

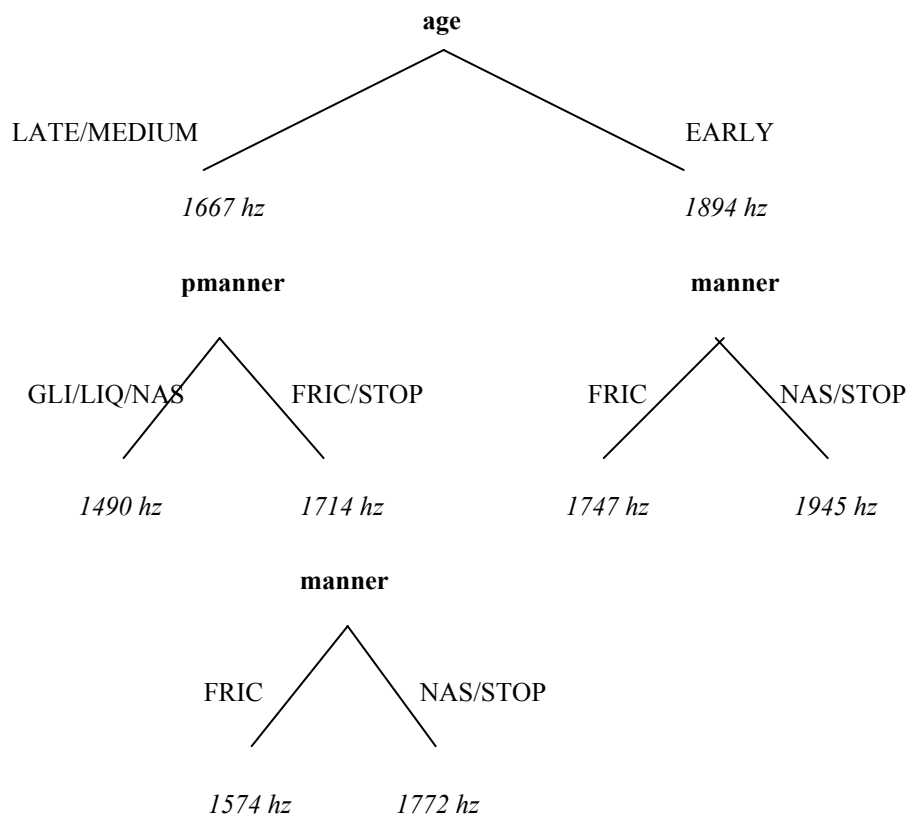


Figure 3.9 (b) – CART analysis of all KIT F2 measurements.

What is somewhat mysterious here is the behaviour of the post-nasals, since for the other two environments, no standard of comparison exists across syllable positions in NZE. That is, glides and liquids do not appear in coda position in NZE, and we therefore do not know how pre-nasals would pattern in comparison to pre-fricatives and pre-stops *if* there existed pre-liquid or pre-glide environments. In the data as presented in figure 3.9 (b), it appears as if nasals projected different coarticulatory patterns onto adjacent vowels depending on whether they precede them or follow them. I have no explanation for this.

Within the DRESS data (figure 3.9 (c)), a low F1 corresponds to a high F2, which constitutes solid evidence of interrelated fronting and raising of DRESS in NZE, and that this pattern is a property of individual systems. Furthermore, we observe allophonic conditioning further down the tree, where tokens following liquids and glides show lower F2 values than those following fricatives, nasals and stops. The fact that this occurs in the left branch only is due to the fact that post-glide/liquid tokens do not occur in the right-branch data at all. Within

the group of tokens with the high F2, pre-velar tokens show higher F2 values than vowels in other environments. In the latter group, pre-nasal vowels are fronter than vowels before fricatives and stops.

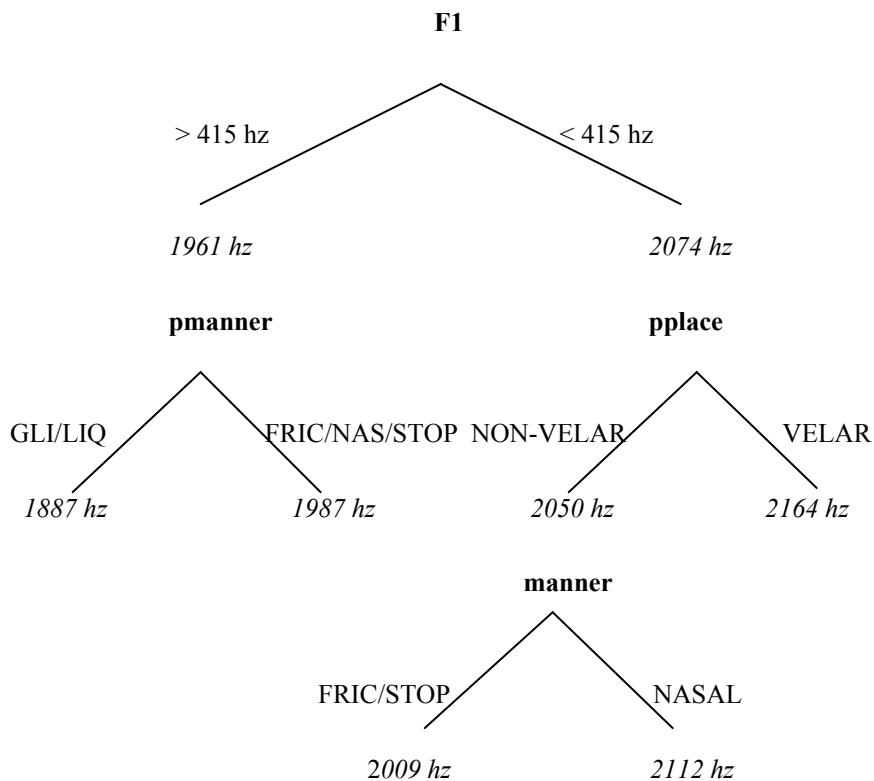


Figure 3.9 (c) – CART analysis of all DRESS F2 measurements.

The TRAP F2 data (figure 3.9 (d)) shows allophonic conditioning at the first branching, with post-palatals and post-dentals showing lower F2 values than post-alveolars/labials/velars. In addition, the former group shows a speaker sex difference with female speakers having higher F2 values than males.

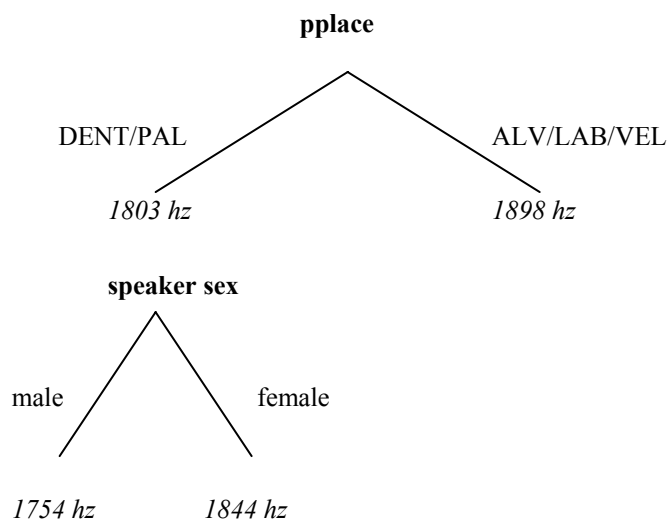


Figure 3.9 (d) – CART analysis of all TRAP F2 measurements.

### *Discussion CART*

The following points emerge from the factor analysis:

- (a) Along most dimensions, allophonic/systemic structure overrides external conditioning by age and/or speaker sex. The exceptions to this are the F2 data of KIT and the F1 data of TRAP, where late and medium speakers show more innovative realisations than early ones. This seems to be due to different reasons: In KIT, the degree of change in F2 over the intermediate period is sufficiently large for the temporally different subsamples to be rather far apart from each other, which increases the robustness of the age distinction. In TRAP, the major reason seems to be the lack of allophonic conditioning in the first place which allows age differences to be recognised by the analysis. This is due to both historical and methodological reasons: the TRAP data is biased toward pre-stop environments, since most pre-fricative as well as pre-nasal environments have gone over to START in ‘broad R’ varieties (this process is discussed at length in chapter 6). Although early intermediate speakers retain flat A in some broad A words, these have been excluded from the current analysis. Although we noted both raising and

fronting in DRESS over the intermediate period in section 3 above, age does not seem to play a major role in predicting frequency values of either formant for that vowel. Rather, the interrelation between the two formants is probably due to it being a property of both the speech community over time as well as individual speakers.

- (b) Both KIT and DRESS show interrelations between the first two formant frequencies. In both cases, F1 and F2 correlate negatively with each other in the CART representations, which indicates that the processes affecting each of the vowels are indeed interrelated. We have furthermore seen that this may hold even at the allophonic level, as the distribution of nasals in one vowel seems to be sensitive to that in the other one (cf. p. 22 above for the structuring of pre-nasal allophones in KIT and DRESS).
- (c) Phonetic allophonisation is, by and large, congruent with what we would expect on the basis of coarticulatory mechanisms. On the whole, there exists a trend in the data to exhibit allophonic conditioning similar to what has been found in laboratory conditions such as Stevens and House's (1963) classical study. More specifically, we have observed that vowels in fricative environments tend to show lower F2s than vowels in nasal and stop-environments. A strong lowering effect is also observed in post-liquid/glide environments. Nasals lead to higher F1 and F2 values on their adjacent vowels with the exception of the late and medium speakers' KIT data, where vowels following nasals shows lower F2 values. It is not entirely clear from a mechanistic point of view why vowels in nasal environments should behave in any way differently from their non-nasal counterparts. After all, the major effect of an adjacent nasal consonant is the introduction of a nasal formant along with dampening of higher formants ('zeros', Fujimura (1962)) with no further influence on the intra-oral articulatory mechanics. However, given the finding that vowels in nasal environments show a smaller perceptual vowel space (cf. Wright 1986), the high F2s we observe in the sample may reflect an articulatory strategy of overshoot to compensate for the decrease in perceptual discriminability. This account, however, is at odds with the increased F1 values found in nasal environments. If anything, we would expect the opposite in the case of high vowels.

- (d) The intermediate period is typologically intermediate with regard to the set-up of SFVs relative to each other. The endpoints of the SFV shifts as sketched in figure 1 represent well the results of studies on early NZE speakers vis-à-vis contemporary ones in that the former tend to have the conservative systems with three front vowels, whereas the latter have a system with two front vowels and one central vowel. Assuming that changes of the kind discussed here come about in a gradual manner, we expect an in-between stage with properties different from either of the endpoints. More specifically, we would expect the centralising vowel (KIT) to traverse the F2 space of TRAP, an expectation that is reflected in the results of the CART analysis above: For pooled data from the entire sample, KIT is grouped together with TRAP in the F2 dimension.

### 3.3.5 Correlations between vowels for individual speakers.

Having identified the primary conditioning factors governing the SFV shift in Intermediate NZE, I will now turn to the question of how both the movements of the three lexical sets as well as their phonemic subsets relate to each other. If the chain-shift hypothesis is correct, we would expect a positive correspondence between F1-values of DRESS and TRAP as well as between F1-values for DRESS and F2-values for KIT. In addition, we can test the hypothesis that the shift came about as a push-chain, which lead to the transitional stage where the lexical sets of DRESS and KIT were close to each other. If this is true, we would expect to find speakers with high DRESS, and uncentralised KIT. Table 3.3 shows Spearman's correlation coefficients between F1 and F2. Figure 3.10 plots mean formant frequency values of each individual speaker in those dimensions where correlations are expected to hold, that is, the F2 value of KIT vs. F1 of DRESS (Fig. 3.10 (a)) and the F1 of DRESS against the F1 for TRAP (Fig. 3.10 (b) and (d)).

Two conclusions can be drawn from the correspondence test results indicated in table 3.3. First, there is a solid positive correlation between the first formant frequency value of DRESS and the F2 of KIT. That is, the lower the F1 of DRESS, the lower the F2 of KIT tends to be. Translated into articulatory terms, the height of DRESS correlates with the centralisation of

KIT. Along similar lines, it is clear that there is a height correlation between the lexical sets of DRESS and TRAP, whereby a low F1 value for DRESS corresponds to a low F1 for TRAP.

KIT <-> DRESS			DRESS <-> TRAP		
OVERALL	<b>cor.</b>	<b>p</b>	OVERALL	<b>cor.</b>	<b>p</b>
KIT F2 <-> DRESS F1	<b>0.49</b>	<b>&lt;.01</b>	DRESS F1<-> TRAP F1	<b>0.75</b>	<b>&lt;.0001</b>
KIT F2 <-> DRESS F2	-0.19	0.31	DRESS F2<-> TRAP F1	<b>-0.44</b>	<b>&lt;.05</b>
KIT F1 <-> DRESS F1	-0.29	0.12	DRESS F2<-> TRAP F2	<b>0.44</b>	<b>&lt;.05</b>
KIT F1 <-> DRESS F2	0.32	0.08	DRESS F1<-> TRAP F1	-0.16	0.41

Table 3.3 – Spearman’s correlation coefficients and their significance levels. The left column gives values for correlations between the lexical sets of KIT and DRESS, the right column those between DRESS and TRAP. The first row states the correlation between the means (for each speaker) of the lexical sets as a whole, the following ones indicate the correlations with respect to the coding-category given in each header row. Significance levels below 0.05 are given in bold.

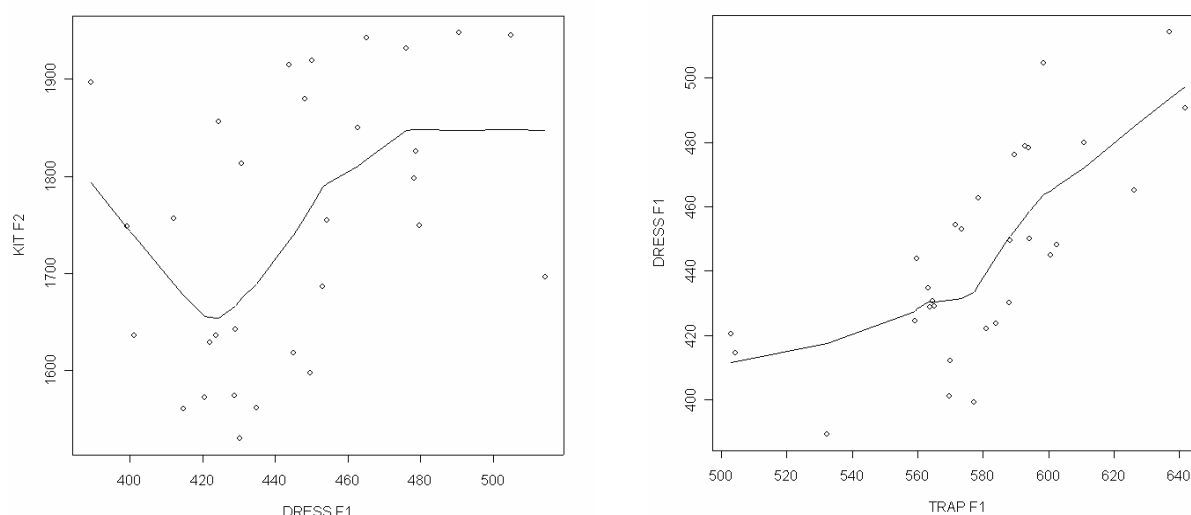


Figure 3.10 - Individual mean values in those dimensions that are assumed to be interrelated (i.e. F2 for KIT vs. F1 for DRESS in figures 3.10 (a) as well as F1 for DRESS vs F1 for TRAP in 3.10 (b). Each point represents mean values for a single speaker. The line represents a non-parametric scatterplot smoother fit through the data (Cleveland 1979).

Therefore, the plots in figure 3.10 provide evidence for the push-chain hypothesis. Recall from table 3.1/figure 3.4 that the EARLY speakers have a set-up where the means of KIT and DRESS are fairly close to each other, which was hypothesised to be indicative of a stage where DRESS had raised, but KIT was still a front vowel. In terms of correspondences, we would then expect there to be individual speakers who have a low mean F1 in DRESS (i.e. a high DRESS vowel) as



well as a relatively high F2 in KIT. This is exactly what we find in figure 3.10 (a), where the upper left corner is populated, whereas the lower right one is not. This becomes clearer once we control for contextual factors. Figure 3.11 below plots speaker averages of DRESS F1 and KIT F2 in pre-alveolar environments only. Whereas the data points are correlated towards higher DRESS F1 values, this correlation is not present toward lower F1s due to the presence of speakers with a high KIT F2 mean and a low DRESS F1 mean. We can therefore conclude that it is a push-shift, and that it is sequential (i.e. the elliptical distribution does not come about by grouping together speakers of different degrees of innovativeness, but represents a ‘true’ transitional stage exemplified in the speech of one and the same speaker).

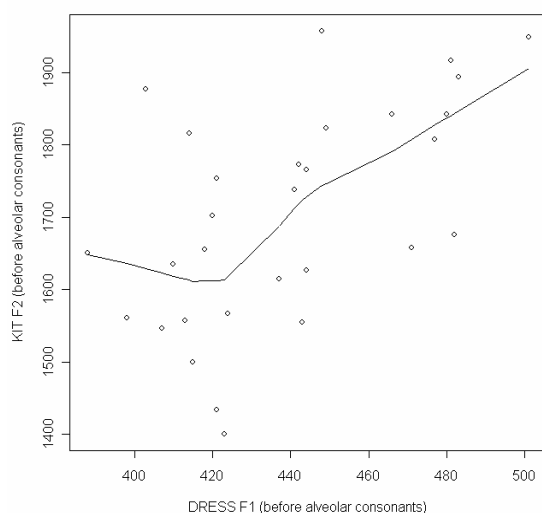


Figure 3.11 - Individual mean values of DRESS F1 and KIT F2 before alveolar consonants. Each point represents mean values for a single speaker. The line represents a non-parametric scatterplot smoother fit through the data.

A more basic point that can be empirically assessed further is the claim that the quality alterations in the three SFVs are indeed partly due to change over time in the first place. Recall that although it seems clear from the data presented in figures 3.1 and 3.2 that change over time occurs over the intermediate period, this factor was to some extent ‘swamped’ by systemic variables such as environment and formant frequency in the regression analysis presented in section 3.5.4. In the remainder of this section I will show results of correlation tests between formant frequency and age (here: year of birth, henceforth yob). Figure 3.12 below plots (for individual mean values) yob against F1/F2 for all three SFVs, and indicates correlation coefficients and their significance levels.

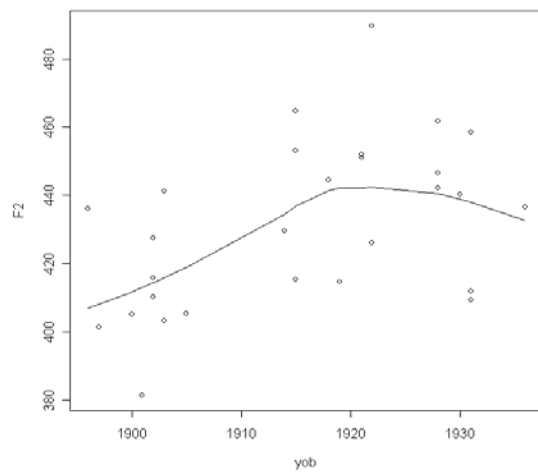
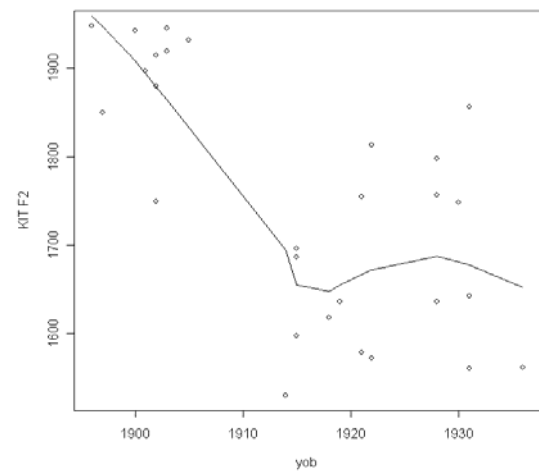
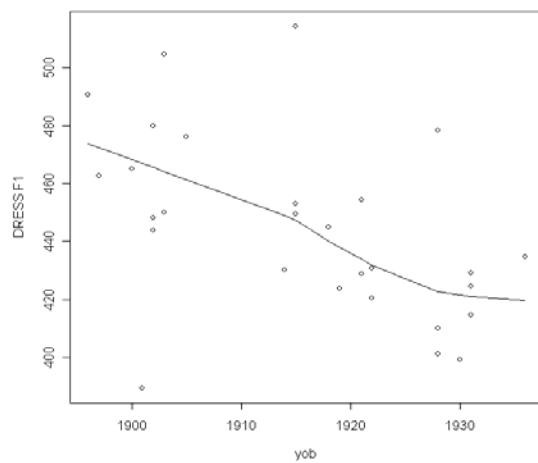
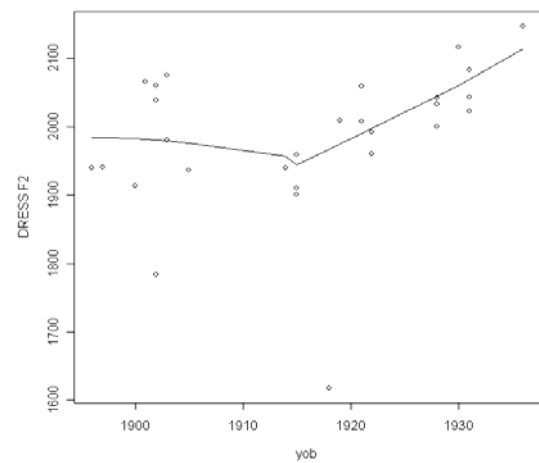
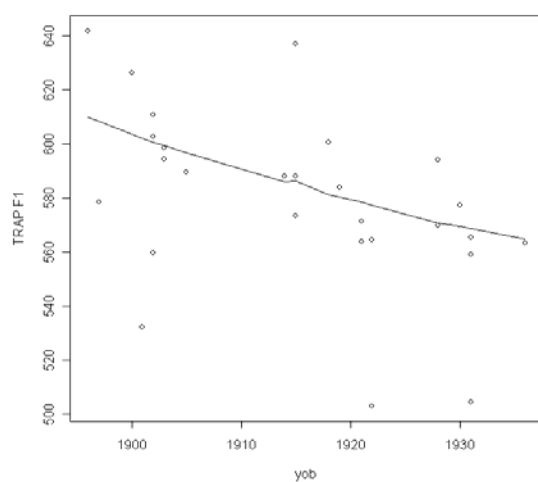
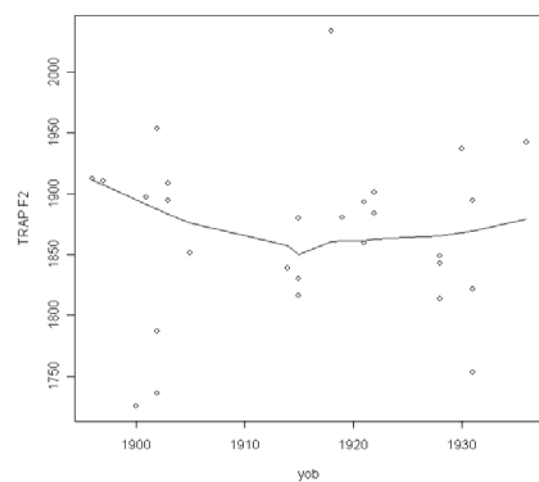
Figure 3.12 (a) –  $Rho = .46$  ;  $p < .05$ Figure 3.12 (b) –  $Rho = -.59$ ;  $p < .01$ Figure 3.12 (c) –  $Rho = -.53$ ;  $p < .01$ Figure 3.12 (d) –  $Rho = .45$ ;  $p < .05$ Figure 3.12 (e) –  $Rho = .56$ ;  $p < .01$ Figure 3.12 (f) –  $Rho = -.03$ ,  $p > .8$ 

Figure 3.12 - Individual mean values of formant frequency and year of birth. Each point represents mean values for a single speaker. The line represents a non-parametric scatterplot smoother fit through the data.

It seems clear that in all dimensions except the F2 dimension of TRAP, there are solid correlations between formant frequency and yob. More specifically, the correlations reflect what we expect on the basis of the plots in section 3 above: In KIT, the correlation between yob and F1 is positive, whereas it is negative between yob and F2, which means that the younger a speaker, the more centralised and lower their realisation of KIT. As for DRESS, the inverse holds: yob correlates positively with F2 and negatively with F1, which is further evidence for raising and fronting of DRESS over the intermediate period. Finally, there is a clear correlation between F1 and yob in TRAP, from which we infer raising of that vowel over the sample. No correlation holds between F2 of TRAP and yob.

However, if we look more closely at the details of the plots in figure 3.11, the picture is slightly more complex: Whereas in the F1 dimension of DRESS and TRAP, the progression toward lower values is approximately linear, this is not the case in the F1 of KIT, nor in the F2 of KIT and DRESS. More specifically, in both formants of KIT, the correlation seems to be carried mainly by the difference between early and non-early speakers, whereas that in the F2 of DRESS rests on the difference between late and non-late speakers. This suggests, over the intermediate period at least, a fine-grained picture of the following type: KIT centralises and lowers as a reaction to nearby DRESS. After that, DRESS fronts while KIT remains stable. The (hypothetical) reasons for this will be discussed in chapter 8 of this thesis in a broader framework, and can be summarised as follows: It makes sense for TRAP and DRESS to raise *in accordo*, since at any time during the process, low F1 values of DRESS are more faithful predictors of DRESS by virtue of being furthest from TRAP. Given that KIT reacts to encroaching DRESS at some initial stage by centralising and lowering, high F2 values become increasingly better cues for DRESS than was the case when KIT had not yet centralised. Hence, the F2 dimension is employed to cue the contrast between DRESS and TRAP as F1s become more similar (note that the crossover in F1s between KIT and DRESS seems to occur approximately at the medium stage of the intermediate period).

### 3.4 Lexical Frequency

One effect that has received growing attention in recent studies on sound change is lexical frequency (Bybee (2001), Bybee and Hopper (2001)). In brief, most of these studies have found an effect whereby high-frequency words show more innovative realisations of a given variable they contain than do rare words, for the following reason: Frequent words are supposed to be more easily retrieved from memory than non-frequent ones, i.e. if they contain the sound undergoing change, the changing quality of that sound is ‘made up for’ by the familiar overall word shape.

In this section we will look at whether there are any frequency effects in the SFV data, and to what extent these interact with structural and external factors as analysed above. That this is a separate analysis is due to two reasons: First of all, lexical frequency was not part of the original analysis of the present study, and is to be regarded as an ‘adjunct’ to it. In addition, it is not entirely clear to what extent lexical frequency effects would be expected to be found in gradual vowel change of the type exemplified by the SFV shift. Most of the studies that have found lexical frequency effects in sound change focus on abrupt types of change such as merger by transfer and change in consonantal quality. In addition, the most convincing cases for lexical frequency effects are those involving the *loss* of some phonemic element, rather than change (cf. the discussion on loss of coda-s in Spanish in Bybee (2001)).

In order to assess frequency effects, we will employ the following measures: (1) CART analysis including lexical frequency as an independent variable, and (2) correlation tests.

#### 3.4.1 Factor analysis

Figure 3.13 (a) shows a CART representation of the KIT F2 data regressed onto the independent variables age, speaker sex, manner, (log) lemma frequency, (log) word frequency. The frequency counts (these as well all frequency counts reported in this thesis) are obtained on the basis of the CELEX lexical database (Baayen et al. (1995)). The trees have been truncated if word frequency did not show up on the third node or lower.

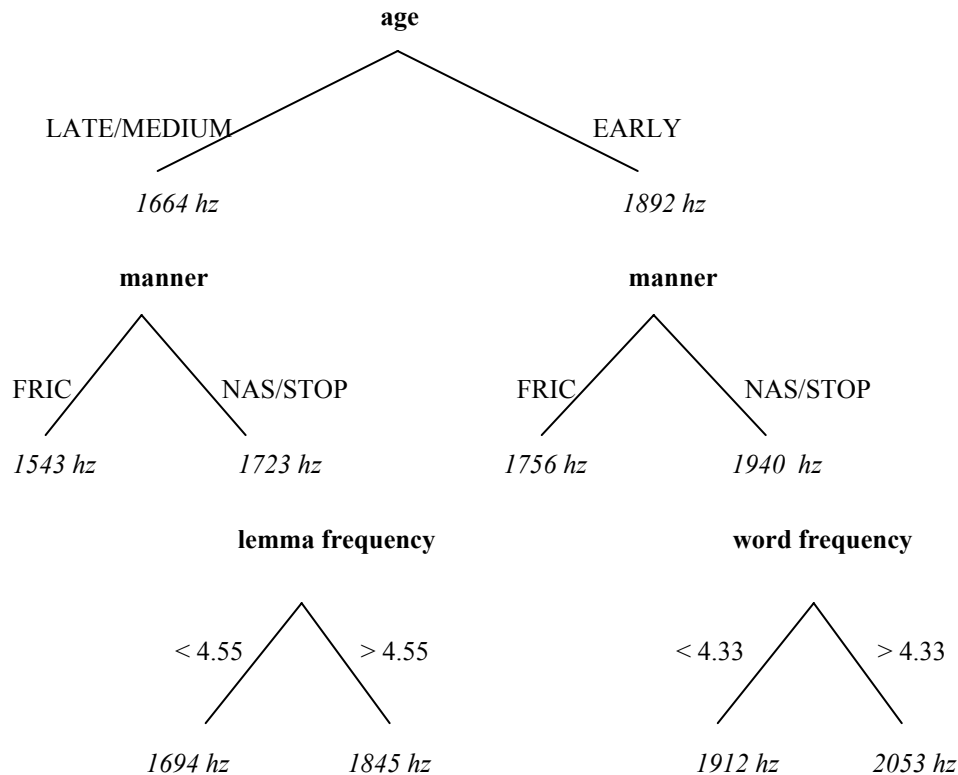


Figure 3.13 – CART analysis of KIT F2 measurements including the independent variables *lemma frequency* (logged) and *word frequency* (logged).

It seems clear that lexical frequency plays a role, if in an unexpected way: For all speaker groups, lexical frequency occurs only within the more conservative allophones of KIT, i.e. those with a high F2 (post-nasals and stops). In addition, the relation is such that vowels occurring in high-frequency items show the more conservative behaviour. Given their occurrence in a conservative superordinate branch, it is vowels in high frequency words that show the most conservative behaviour overall.

Lexical frequency does not show up in any other one of the dependent variables in the sample within the first three branching points. This, however, may be due to other factors simply overriding minor effects of lexical frequency, and it might be useful to consider lexical frequency vs. formant frequency exclusively. Note that if we posit an effect of lexical frequency onto innovativeness of a given token, and that innovativeness is mainly determined by formant frequency, we would expect correlations between lexical frequency and formant frequency. Table 3.4 shows results for a Spearman correlation test of F1/F2 vs. both lemma and word frequency.

		<b>frequency</b>			
		<b>lemma</b>		<b>word</b>	
		<i>rho</i>	<i>p</i>	<i>rho</i>	<i>p</i>
<b><i>KIT</i></b>	F1	-0.04	>.1	-0.02	>.43
	F2	<b>0.15</b>	<b>&lt;.0001</b>	<b>0.15</b>	<b>&lt;.0001</b>
<b><i>DRESS</i></b>	F1	<b>0.06</b>	<b>&lt;.01</b>	0.06	>.1
	F2	<b>0.06</b>	<b>&lt;.05</b>	<i>0.04</i>	>.09
<b><i>TRAP</i></b>	F1	-0.01	>.5	0.009	>.07
	F2	<b>-0.08</b>	<b>&lt;.01</b>	<b>-0.13</b>	<b>&lt;.0001</b>

Table 3.4 – Correlation coefficients and significance levels between word frequency/lemma frequency and F1/F2 for all three SFVs.

Table 3.4 shows rather solid correlations between lexical frequency and formant frequency in all vowels and most dimensions. The exact relations, however, are not straightforward: Whereas high frequency tokens correlate positively with innovativeness in the F2 dimension of DRESS, the opposite holds true in the F2 of KIT and TRAP as well as the F1 for DRESS.

These findings seem to contradict reports on lexical frequency effects in the intermediate period (Hay and Bresnan (2006)), where positive correlations between frequency and innovativeness were found. However, the analysis given above does not discriminate between content words and function words. It has been found that these differ substantially in the degree to which they drive innovation in a given sound they contain, and that various lexical frequency effects may in fact be an artefact of this pattern (i.e. a given function word is more frequent than a given content word, hence, if it is occurrence in a function word that is the ‘true’ reason for innovativeness, the lexical frequency effect occurs automatically).

The data reviewed in the current paper was not initially coded for whether the word was a function word or a content word, which means that an investigation of this problem must remain tentative in this regard. If we restrict the correlation analysis to tokens that are clearly content words (i.e. leaving out function words and indeterminate cases), we obtain the numbers as stated in table 3.5 below:

		frequency			
		lemma		word	
		<i>rho</i>	<i>p</i>	<i>rho</i>	<i>p</i>
<b>KIT</b>	F1	-.05	> .1	-.02	> .4
	F2	<b>.09</b>	<b>&lt; .001</b>	<b>.08</b>	<b>&lt; .01</b>
<b>DRESS</b>	F1	<b>.06</b>	<b>&lt; .05</b>	<b>.05</b>	<b>&lt; .05</b>
	F2	<b>.07</b>	<b>&lt; .01</b>	<b>.05</b>	<b>&lt; .05</b>
<b>TRAP</b>	F1	.01	> .7	-.0001	> .9
	F2	.05	> .1	.03	> .4

Table 3.5 - Correlation coefficients and significance levels between word frequency/lemma frequency and F1/F2 for all three SFVs; excluding function words.

We note that essentially the same relation hold in comparison to table 3.4, if at somewhat higher p-values. The frequency effect in TRAP does not occur in the content words. This shows that the observed frequency effects are not an artefact of the special behaviour of function words, as these seem to, if anything, magnify the effect (i.e. lead to lower p-values in table 3.5). For example, the F1/F2 values of DRESS function words are 471hz/1880hz compared to 454 hz/2000hz for content words, which is too say that the function words are rather more conservative than the non-function words. However, all instances of the former group are preceded by /w/, which in all likelihood would be expected to bring about lowering of adjacent formant values (but see Moon and Lindblom (1994) for an alternative scenario). In addition, the fact that the function words appear to be rather more conservative than the content words appears to also be due to their distribution being biased, in that more function words appear in the early sample. For the words *in* and *if* (the most frequent function words in KIT), we obtain the following token numbers:

	Early	Medium	Late
if	6	3	3
in	25	3	7

Table 3.6 – Token numbers of the words *if* and *in*, broken up into subsamples.

On the basis of the current data it might therefore be advisable to remain cautious about the (here: inhibitive) role of lexical frequency in this particular sound change.

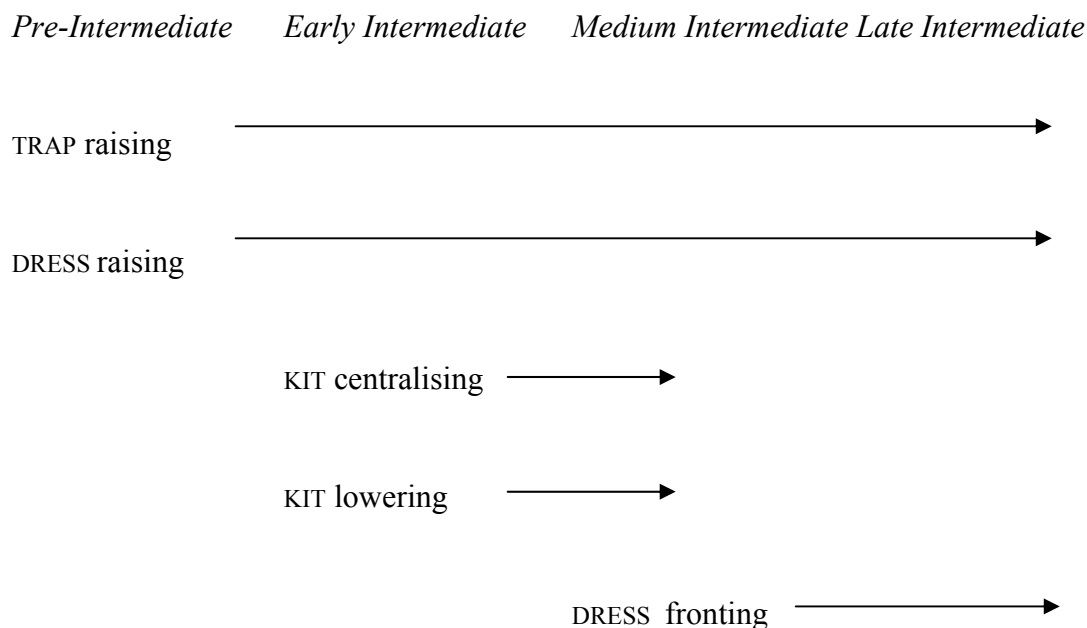
### 3.5 Discussion – SFVs in F1/F2 space

In the above analysis of the SFVs in F1/F2 space it was shown that the Intermediate Period sees a change in the typology of the short front vowels, whereby an earlier system of 3 front vowels changes into a set-up with two front vowels, DRESS and TRAP, and a central vowel, KIT. Furthermore, it could be observed that the behaviour of the KIT vowel is less straightforward than the movement of the overall means over time suggest. Rather, different allophonic categories shift at different rates, and centralisation is never into empty territory *sensu stricto*, since a more centralised mean in a later group of speakers is occupied by one or more innovative allophones in the speech of an earlier group. In addition, there is one group of speakers (the EARLY MALES) that shows a most pronounced stretch in the F2 dimension of the KIT vowel, which suggests the possibility of a temporary system with both fronted as well as centralised allophones of KIT.

The CART analysis revealed that the distribution of F-values is predicted mainly by phonological environment, and that the nature of structural conditioning is mostly in accordance with what we would expect on coarticulatory grounds.

In addition, a number of correlation test results were discussed. It could be shown that the historical process that converted a system of three short front vowels into one of two front vowels and a central vowel is indeed a chain-shift. In addition, it was shown that the overall process may be describable in terms of temporarily somewhat disjunct subprocesses, since correlations between age and formant frequency are not strictly linear in all instances. Based on this, we can infer the following chronology of the SFV shift from the MU to the late Intermediate stage:





### 3.6 Vowel duration

The above analysis has focussed exclusively on the formant frequency dimension, and the patterns and mechanisms involved in the SFV shift have been explained as interrelations between the three vowels in that dimension. This continues the traditional approach to studying the SFV shift in NZE, which has also focussed on vowel quality as the primary, or even the exclusive, explanandum in the restructuring of these vowels. We do, however, have reasons to believe that vowel *quantity* (i.e. duration) has had a role to play in the restructuring of the NZE front vowel space on the grounds that a number of recent studies (Maclagan and Hay (2006), McKenzie (2005)) have shown that duration (as well as diphthongisation) are important cues for vowel identity in high front vowels. More specifically, it was pointed out that DRESS is now a comparatively long vowel, which we would not expect if the overall development involved only quality changes under etymologically faithful continuation of the quantitative properties of these vowels, all of which originate as short vowels with some corresponding long vowel in Middle English.

Given this asymmetry, it seems reasonable to infer change in the quantity dimension as well. This section will therefore present results obtained from duration measurements of a

subsample of speakers (the early male and the late female speakers, which represent the most conservative and the most innovative stages, respectively). Section 3.8.1 summarises results from a number of studies on vowel duration. Section 3.8.2 presents an analysis of the SFVs in the intermediate sample. 3.8.3 discusses potential implications of the findings. 3.8.4 relates the quantity dimension to the quality dimension and advocates an integrated view on vowel spaces.

### 3.6.1 Background – Vowel duration in English

There exists the persistent myth that vowel length is not distinctive in English (cf. Pinker 1995: 168). This has to do with the fact that certain minimal pairs (such as KIT and FLEECE) are distinguished by both duration (FLEECE being longer) as well as quality (FLEECE being more peripheral in the sense of Labov 1994).<sup>25</sup> Similar relations hold in most parts of the vowel space (cf. START vs. STRUT, THOUGHT vs. LOT, FOOT vs. GOOSE). Whereas in ‘true’ quantity systems such as Finnish, vowels are distinguished on the basis of length alone, Germanic languages have adopted this ‘dual encoding’ of adjacent vowels. However, once we move away from how these vowels are traditionally represented in phonemic transcriptions (which uses different symbols for let’s say the vowels in START and STRUT) and from RP/GenAm, it becomes clear that duration does indeed play a role in the phonemic classification of a number of vowels. First of all, a number of studies on American English have shown that there exist solid durational differences between a number of vowels of the same height (Peterson and Lehiste (1960), Crystal and House (1982/1988 (a),(b))).

In addition, AusE as well as NZE have a number of vowels virtually on top of each other in formant space, especially START and STRUT as well as KIT and FLEECE (in AusE) and DRESS and FLEECE (in NZE, cf. Maclagan and Hay (2006)). These vowels, however, are

---

<sup>25</sup> It should be pointed out that the remarks by Pinker seem to relate to the inappropriate classification of English vowels into long and short on the basis of a pre-GVS system by high-school teachers (where ‘long vowels’ are in fact the diphthongs of modern English). Although this is indeed wrong from a synchronic point of view, the conclusion that vowel length is not significant in present-day English is not warranted as new long/short pairs have arisen which are, however, additionally distinguished by quality differences. These qualitative differences have, in turn, been analysed as constituting the primary cue toward distinguishing the long/short pairs (cf. Bauer 1994).

clearly not merged, which seems to be at least partly due to solid length differences between them (*ibid.*)<sup>26</sup>.

This typology of subsystems has led to a reconsideration of terminology in order to capture both the fact that there is ‘duality in cueing’ in pairs such as FLEECE/KIT as well as the fact that the two subsystems tend to show different behaviour in historical sound change. The relevant terminology is laid out in Labov (1994), who classifies vowels of the FLEECE type as ‘tense’ vowels and vowels of the KIT type as ‘lax’ vowels. Labov goes on to show that tense vowels tend to rise in the history of a number of languages (such as German, English, Dutch and Swedish), whereas lax vowels tend to lower. Given the classification of DRESS and TRAP as ‘lax’, the Southern hemisphere varieties seem to pose an interesting counterexample to the general principles advocated by Labov (This will be discussed at length in chapters 7/8).

Coming back to actual empirical facts, the duration ratios given in table 3.7 below emerge from a number of studies on phonetic vowel length in American English (The numbers given in table one are averages emerging from these studies, cf. Peterson and Lehiste (1960), Port (1981), Crystal and House (1982/1987/1988 (a),(b)), Luce and Luce (1985)). The baseline is set at 1 for the shortest vowel, KIT before voiceless stops. For example, the THOUGHT vowel before voiceless stop is 1.72 times longer than KIT in the same environment if we average over the results from the studies mentioned above.

		before voiceless stops	before voiced stops
short vowels	KIT	1	1.4
	DRESS	1.1	1.54
	STRUT	n/a	n/a
	FOOT	1.1	1.54
long vowels	FLEECE	1.4	1.96
	TRAP	1.83	2.56
	START	n/a	n/a
	THOUGHT	1.72	2.41
	GOOSE	1.44	2.01
	NURSE	n/a	n/a

Table 3.7 – Duration ratios between different vowels before voiced and voiceless stops in American English. Note that the START/STRUT distinction was not consistently held across all the studies from which these numbers were calculated. Further note that no LOT vowel was listed in any of the above-quoted studies, which suggests that the subjects (or the researchers) had the LOT-THOUGHT merger. The NURSE vowel was also not mentioned in all the studies, which is presumably due to its r-colouring in American English.

<sup>26</sup> ‘partly’ because it may be hypothesised that the FLEECE/DRESS (in NZE) and the FLEECE/KIT (in AusE) distinction may be carried in part by the greater degree of diphthongisation of the former vowel.

The following generalisations seem to hold: The vowels in FLEECE/TRAP/(START)/THOUGHT/GOOSE/NURSE contrast with the ones in KIT/DRESS/(STRUT)/FOOT in that the former group has significantly higher duration on average. In addition, there exists a correlation between length and openness within each of the two groups, open vowels have been shown to be longer than close vowels (cf. Peterson and Lehiste (1960)). Some further points that have emerged from the above studies deserve mentioning in this context: Although the evidence is not uniform, vowels in monosyllabic words are longer than vowels in polysyllabic words (cf. Port 1981, Klatt 1973, although there seems to be counterevidence; cf. Umeda 1975). This effect does not only affect vowel duration as a whole, but also the length differential between vowels depending on whether they occur before voiced or voiceless stops. (Klatt (1973) found that in monosyllabic words, vowels before voiced stops are 34% longer than before voiceless stops. In polysyllabic words, this difference reduces to 22%. From this, Klatt infers that various shortening effects such as closeness/following voiceless stop/laxness do not stack up linearly, but rather ‘smooth out’ (if plotted on a curve) as lower durations are reached).

Although there does not seem to be a study of comparable magnitude on vowel length in NZE, researchers on that variety agree on a similar grouping into long and short vowels, with the exception of TRAP, which is commonly referred to as a short vowel in NZE.

What is more controversial is the role of openness vis-à-vis length, in that the solid correlation observed for American English does not seem to be reflected in present-day NZE (cf. Gordon et al. (2004: 270); who furthermore note that this correlation was indeed present in the Mobile Unit data). I will show in this chapter that the correlation of openness with duration is basically non-existent in the data analysed in the present context. Although DRESS and TRAP do indeed rise over the Intermediate Period, they maintain their duration.

### 3.6.2 Vowel duration in the Intermediate period

In what follows, I will present an analysis of duration measurements of a restricted number of speakers from the original sample that formed the basis of the formant analysis presented

above. Only the two ‘extreme’ groups were analysed, i.e. the EARLY MALE group and the LATE FEMALE group. In order to control for contextual factors, only pre-alveolar pre-plosive tokens were selected for DRESS and TRAP, and pre-alveolar as well as pre-velar plosives for KIT (since they constitute the extreme ends of the KIT distribution along the F2 axis). I attempted to obtain 10 pre-alveolar tokens for each speaker and each vowel as well as 10 pre-velar KIT vowels for each speaker. However, this was not always possible. The token numbers obtained from each speaker/vowel are given in table 3.8. I furthermore attempted to obtain only monosyllabic tokens (due to the pervasive duration difference in vowels depending on the number of syllables the word they occur in has, cf. section 3.6.3 above), which was not possible for each individual speaker. The number of monosyllabic as well as polysyllabic tokens obtained from each speaker is summarised in table 3.9. In the analysis below, we will therefore consider monosyllables and polysyllables separately.

		KIT		DRESS		TRAP	
Group	Speaker	_[+voice]	_-[-voice]	_[+voice]	_-[-voice]	_[+voice]	_-[-voice]
Early	1	17	22	6	2	7	5
	2	7	17	4	4	5	6
	3	9	17	7	6	10	6
	4	11	6	5	3	6	5
	All	44	62	22	15	28	22
Late	1	7	9	5	4	6	5
	2	5	6	3	2	3	5
	3	8	5	5	7	5	6
	4	5	7	1	1	1	5
	5	4	3	0	1	3	5
	All	29	30	14	15	18	26

Table 3.8 – Number of tokens obtained from two groups of Intermediate speakers. [+/-voice] specifies the voicing state of the following consonant.

Group	Vowel	Syllabicity <i>mono</i>	<i>poly</i>	<i>ratio</i>
Early	<b>KIT</b>	61	45	1.35
	<b>DRESS</b>	24	13	1.84
	<b>TRAP</b>	32	18	1.77
Late	<b>KIT</b>	32	27	1.18
	<b>DRESS</b>	21	8	2.62
	<b>TRAP</b>	34	10	3.4

Table 3.9 – Token numbers obtained from two groups of Intermediate speakers broken down into occurrences in monosyllabic and polysyllabic words.

The data obtained from the 9 intermediate speakers are shown in table 3.10:

	<i>monosyllables</i>			<i>polysyllables</i>		
	<b>KIT</b>	<b>DRESS</b>	<b>TRAP</b>	<b>KIT</b>	<b>DRESS</b>	<b>TRAP</b>
<i>All</i>	49.65	99.37	106.94	46.64	75.57	93.46
<i>Early</i>	51.66	104.1	103	49.27	74.08	94.44
<i>Late</i>	45.81	94.38	110.65	42.43	78	91.7
<i>_[+voice]</i>	59.23	115.3	118.7	46.32	76	97.75
<i>_-[-voice]</i>	42.72	72.44	97.17	46.92	73.29	87.75
<i>Early _[+voice]</i>	62.58	114.4	112.1	48.39	77.33 (n=3)	96.73
<i>Late _[+voice]</i>	52.54	116.9	127.2	44	82.25 (n=4)	100
<i>Early _[-voice]</i>	43.54	69.2	92.67	49.85	73.1	90.86
<i>Late _[-voice]</i>	42.21	73.91	100.4	40.33	73.75	83.4 (n=4)

Table 3.10 – Average durations in milliseconds (ms) of short front vowels in the speech of 9 intermediate speakers, 4 early males and 5 late females. Token numbers below n=5 are indicated.

### 3.6.2.1 Vowels in Monosyllables

The data given in table 10 suggest a number of pertinent points. First of all, as can be seen in the second column, both DRESS and TRAP are appreciably longer than KIT in the speech of the Intermediate speakers, where the overall ratios are  $\sim 1/2$  for KIT vs. DRESS,  $\sim 1/2.1$  for KIT vs. TRAP and  $1/1.05$  for DRESS vs. TRAP in monosyllables. The duration difference between KIT and DRESS is significant (Wilcoxon test,  $W = 442.5$   $p < .0001$ ), the difference between DRESS and TRAP is not. The same pattern holds if we break down the data into the two age groups. The differences between the age groups are not significant. In addition, the well-known length difference within a vowel depending on whether it occurs before a voiced stop or a voiceless stop holds within the present sample (where the difference in the KIT vowel is significant at  $p < .0001$  ( $W = 1607.5$ ), at  $p < .0001$  for DRESS ( $W = 385.5$ ) and at  $p < .01$  for TRAP ( $W = 776.5$ )). If the data is broken down further into age+voicing of the following consonant, the duration difference between KIT vowels before voiced and voiceless consonants fails to

reach significance in the late sample. The differences in DRESS and TRAP retain significance within both groups.

Overall, this accords with the internal partitioning within the SFV set with respect to the chain-shift in F1/F2 space, in that it is DRESS and TRAP that underwent an upward movement, whereas KIT centralised. Here we have the same sub-grouping (KIT vs. DRESS and TRAP).

### 3.6.2.2 Vowels in Polysyllables

If we compare the durations of the SFVs in monosyllables to those in polysyllables, it seems clear that whereas DRESS and TRAP are shorter in polysyllables, this does not hold for KIT. The duration difference between monosyllables and polysyllables is significant for DRESS only ( $W = 276.5$ ,  $p < .01$ ). In addition, it is only in polysyllables where the duration difference between DRESS and TRAP is significant ( $W = 153.5$ ,  $p < .01$ ). The contrast in duration between pre-voiced and pre-voiceless tokens fails to reach significance in all vowels, which might be partly due to low token numbers.

It seems that the results quoted above as to the effect of polysyllabicity on duration have some evidence in the present analysis (i.e. the DRESS vowel). In addition, there is some support for Klatt's (1973) assumption that there is no linear 'adding up' of shortening factors but rather some minimal level of vowel duration around which shortening factors do not apply anymore. This is exemplified by the KIT vowel, where duration differences are minor and reach significance only between two conditions: Early male monosyllables closed by a voiced stop vs. Early male monosyllables closed by a voiceless stop ( $W = 733$ ,  $p < .0001$ ) as well as Early male monosyllables closed by a voiced stop vs. Early male polysyllables closed by a voiced stop ( $W = 348.5$ ,  $p < .01$ ).

On the whole it seems that shortening factors do indeed interact. Figure 3.14 shows duration averages of all three vowels before voiced and voiceless stops in monosyllables (Fig. 3.14 (a) and (b)) and polysyllables (fig. 3.14 (c) – (d)). It seems as if for a vowel to be subject to shortening factors, it needs to have a certain minimum length. That is, the only significant shortenings occur in DRESS/TRAP where monosyllabic tokens before voiced stops undergo

similar degrees of shortening in both polysyllables as well as in voiceless environments. In addition, these shortening factors do not stack. The KIT vowel is, by and large, exempt from either process with the exception of Early male KIT before voiced stops, which undergoes shortening.

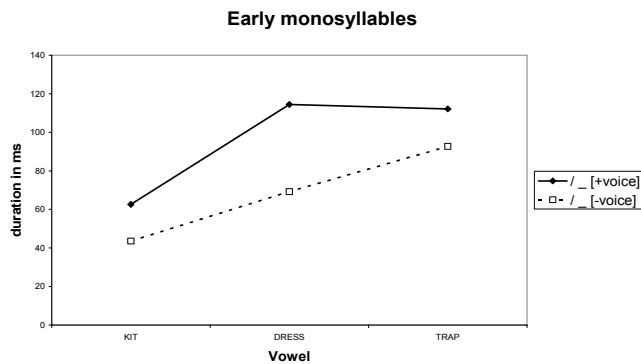


fig. 3.14 (a)

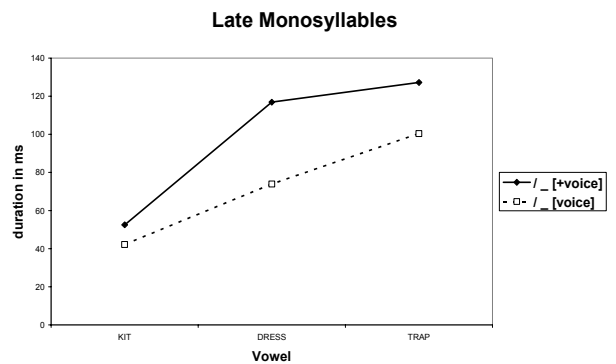


fig. 3.14 (b)

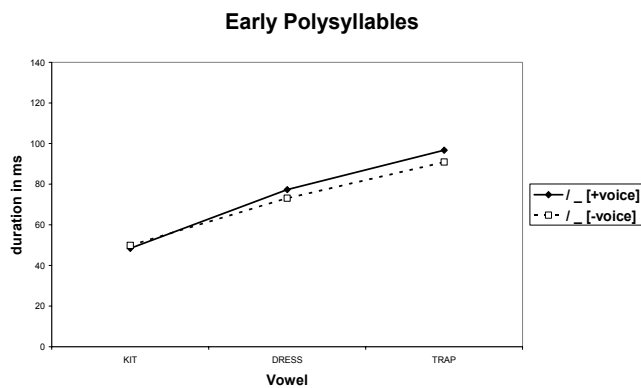


fig. 3.14(c)

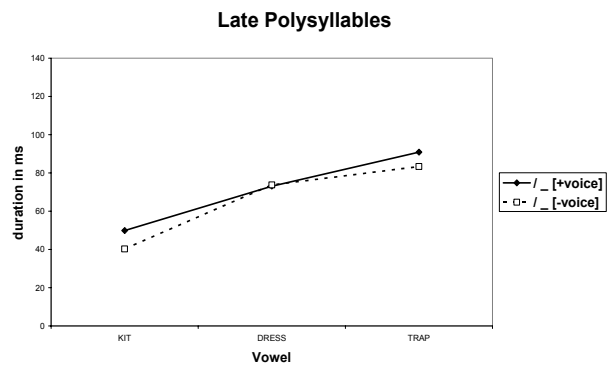


fig. 3.14 (d)

Figure 3.14 – Average durations of KIT, DRESS and TRAP before voiced and voiceless stops. 3.14 (a) shows monosyllables in the early male sample, 3.14 (b) monosyllables in the early female sample. 3.14 (c) and (d) show the same vowels in polysyllables for the same speakers.

As far as duration averages are concerned, we conclude that DRESS and TRAP are not only significantly longer than KIT, but also have a number of properties typical of long vowels such as susceptibility to shortening in certain environments as well as a greater internal (i.e. within one vowel) durational variability.



### 3.6.3 Factor analysis

Figure 3.15 below shows the result of a regression analysis carried out by taking into account a further continuous factor, namely formant frequency. In figure 3.15 duration is the dependent variable regressed onto the following independent variables: Vowel (*KIT/DRESS/TRAP*), F1, F2, Age (*EARLY/LATE*), Syllable (*mono(syllabic)/poly(syllabic)*), Voice (*voiced/voiceless*, i.e. of the following consonant).

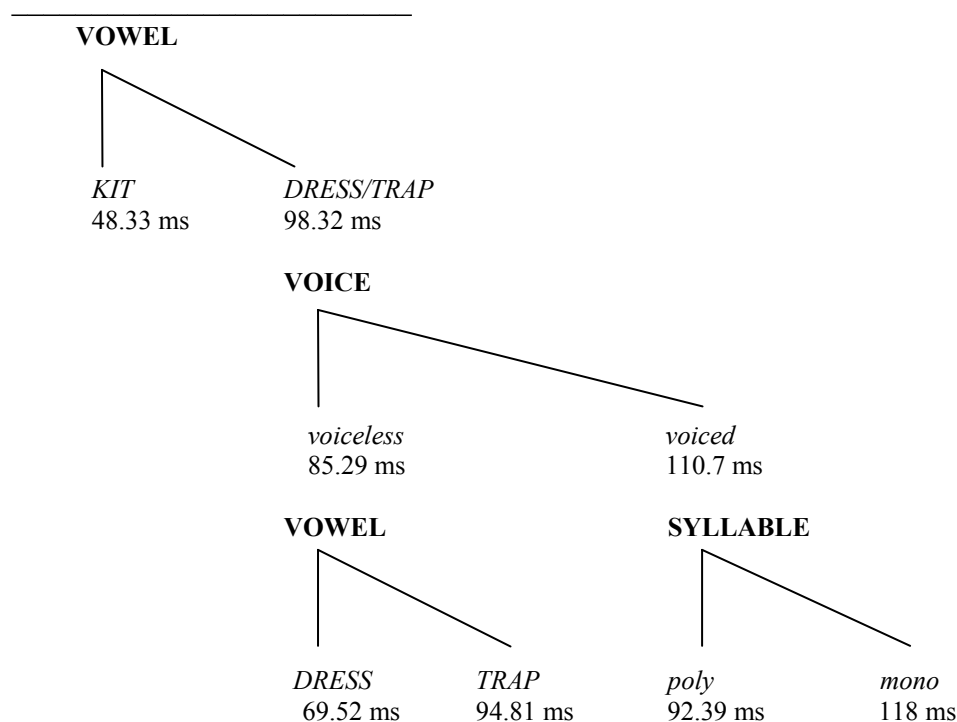


Figure 3.15 – Classification and Regression Tree of pooled duration measurements along the independent variables F1/F2/age/vowel/voice/syllable

As was to be expected on the grounds of the data presented before, the main predictor of overall duration is vowel identity, whereby KIT is appreciably shorter than DRESS and TRAP, respectively. In addition, no further factors partition the KIT branch whereas the DRESS/TRAP branch is split into vowels before voiceless stops (average duration 85.29 ms) and voiced stops (average duration 110.7 ms). This distinction takes precedence over vowel identity, which is a factor only within the *voiceless* node, where DRESS is shorter than TRAP. For DRESS/TRAP vowels before voiced stops it is syllable structure which predicts duration in that both DRESS and TRAP tend to be markedly longer if they occur in monosyllables closed by a

voiced stop than is the case in polysyllables closed by a voiced stop. The two major conclusions from section 3.6.2 therefore seem reflected in the pooled CART data as well: ‘Shortness’ of a vowel goes along with a lack of further internal partitioning along the duration dimension (as KIT represents a terminal node, whereas DRESS/TRAP do not) and shortening factors interact, although with respect to fig. 3.15 above, we may want to prefer to say ‘lengthening factors interact’ since the terminal branching which relates to a phonetic lengthening factor (i.e. syllabicity) stems from the superordinate node which has the higher duration average (i.e. the DRESS/TRAP tokens before voiced stops). Before voiceless stops, vowel identity is a better predictor of duration (where DRESS is shorter than TRAP).

What is striking is that in the pooled data formant frequency does not show up as a factor in predicting vowel length, although this would be expected on the grounds that F1 has frequently been found to correlate positively with duration (cf. the references in section 3.6.1 above). In addition, ‘peripherality’ has previously been related to length (cf. Labov, Steiner and Yaeger (1972), Labov (1994) for a discussion of the relevant terminology) to the effect that the more peripheral a vowel, the longer it would be expected to be (it has to be borne in mind though that the above-quoted authors are concerned mainly with length as a phonological concept and not necessarily phonetic duration). Since we are concerned with front vowels, we would therefore expect a positive correlation between F2 and duration.

This interaction does indeed show up if we break up the overall data pool into vowels. Figure 3.16 (a) – (c) shows CART trees for all three vowels. The variables are as above. The trees are truncated below the second branching.

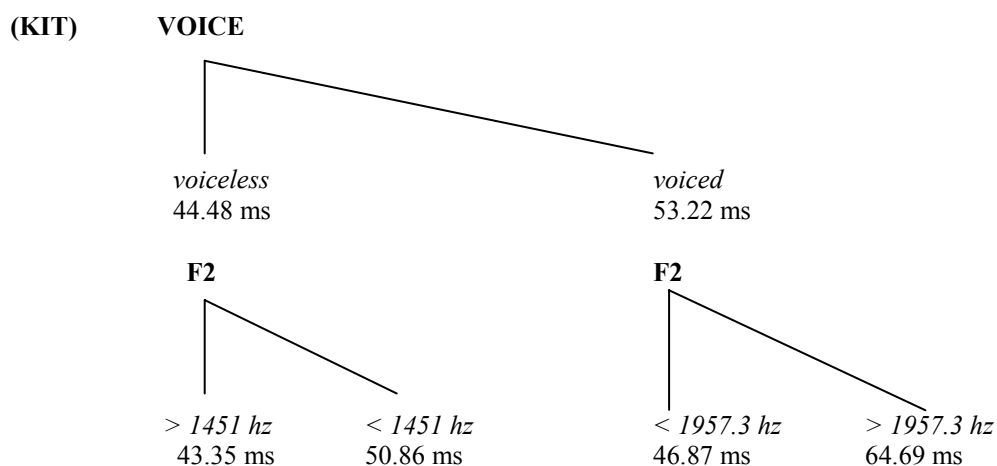


Fig. 3.16 (a)

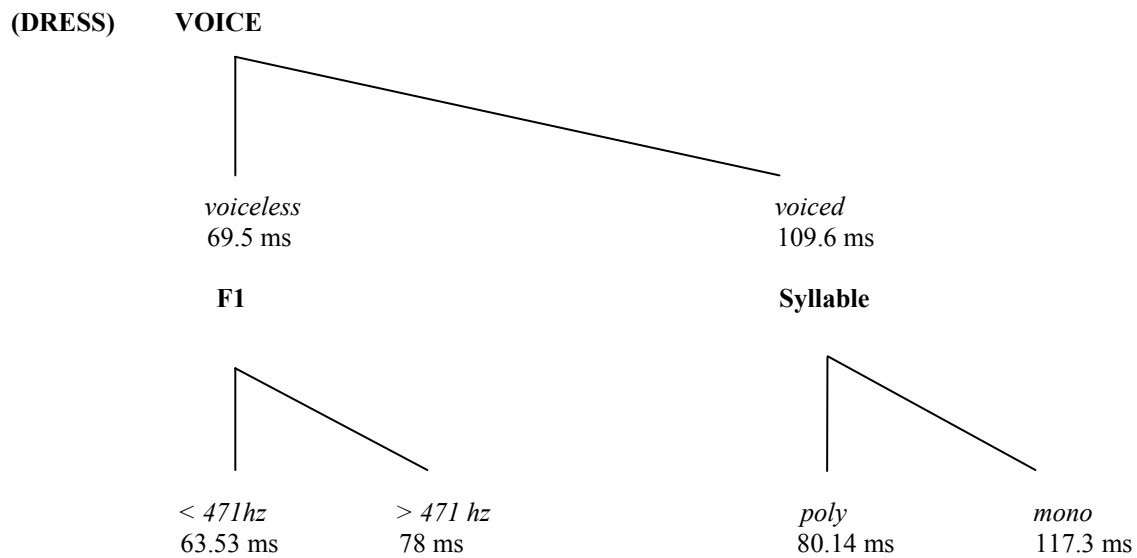


Fig. 3.16 (b)

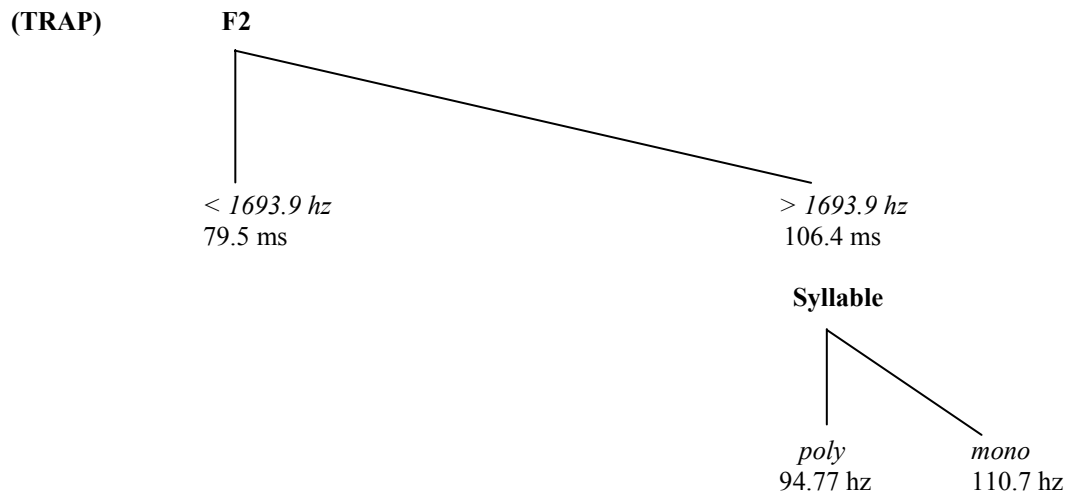


Fig. 3.16 (c)

Figure 3.16 – Classification and Regression Tree analysis of duration measurements broken down into the three SFVs.

Within both the KIT and the DRESS branch, the major predictive factor in determining duration is the voicing status of the following consonant: Vowels before voiced consonants are longer than before voiceless ones (where the ratios are 1/1.2 for KIT and 1/1.58 for DRESS). Two points deserve mentioning in this context: First of all, it is somewhat surprising to find this division in the KIT data as it seems to be mainly a property of the early group, which implies

that if anything, we would expect this as a subordinate branching under a superordinate branching along the factor *AGE*. On the other hand, the late group does show a length difference (albeit not significant) within KIT to the same effect, which might solidify the trend in the pooled KIT data. Secondly, it looks as if the duration difference of vowels before voiced vs. voiceless consonants which we noted as a property of the overall DRESS/TRAP data in figure 3.16 above was carried mainly by DRESS, as it fails to show up as a predictive factor in the TRAP data (3.16 (c)). In the KIT data, second formant frequency is a subordinate factor below both primary branches. The data is somewhat contradictory since within the group of tokens before voiceless consonants, high F2 values seem to predict lower vowel duration, whereas within the sample of vowels which occurs after voiced stops, the inverse is the case. Although this seems unexpected, the vowel tokens for which the negative correlation holds constitute the shortest sub-sample overall, and if we recall the notion as outlined above that the longer the intrinsic duration of a vowel, the more well-in-place the phonetic mechanisms which favour lengthenings/shortenings, this is probably the sub-sample where overriding of the expected mechanisms is least unexpected.

The TRAP data is best predicted by second formant frequency in that the higher the F2, the longer the duration of a vowel token. In addition, we note in both the DRESS and the TRAP data an effect of syllable structure whereby vowels in monosyllables have higher duration averages than vowels in polysyllables within a branching that is already of comparatively high average duration.

It is interesting to note that the expected correlation between F1 and height shows up only within one terminal branch, and for one vowel (DRESS) only. In addition, it fails to reach significance in a Spearman type correlation test ( $Rho = .27$ ,  $S = 2955$ ,  $p > .1$ ). Overall correlation test results are summed up in table 3.11.

F1				F2			
		DURATION				DURATION	
		Rho	p			Rho	p
<i>KIT</i>	all	-0.11	>.1	<i>KIT</i>	all	0.12	>.1
	early	-0.11	>.3		early	0.1	>.2
	late	0.03	>.7		late	0	>.9
	mono	-0.14	>.1		mono	<b>0.2</b>	<b>&lt;.5</b>
	poly	-0.06	>.6		poly	0	>.9
	/ _[+voice]	-0.18	>.1		/ _[+voice]	<b>0.39</b>	<b>&lt;.001</b>
	/ _[-voice]	0.1	>.3		/ _[-voice]	-0.13	>.2
<i>DRESS</i>	all	-0.16	>.1	<i>DRESS</i>	all	<b>0.34</b>	<b>&lt;.01</b>
	early	-0.22	>.1		early	<b>0.46</b>	<b>&lt;.01</b>
	late	-0.18	>.3		late	0.21	>.1
	mono	-0.23	>.1		mono	<b>0.38</b>	<b>&lt;.05</b>
	poly	0.25	>.2		poly	-0.03	>.8
	/ _[+voice]	-0.09	>.6		/ _[+voice]	0.26	>.1
	/ _[-voice]	0.27	>.1		/ _[-voice]	-0.22	>.2
<i>TRAP</i>	all	-0.03	>.7	<i>TRAP</i>	all	<b>0.24</b>	<b>&lt;.05</b>
	early	0.06	>.6		early	-0.01	>.9
	late	-0.04	>.7		late	<b>0.54</b>	<b>&lt;.001</b>
	mono	-0.02	>.8		mono	<b>0.39</b>	<b>&lt;.01</b>
	poly	0.03	>.8		poly	-0.19	>.3
	/ _[+voice]	0.03	>.8		/ _[+voice]	0.13	>.3
	/ _[-voice]	0.06	>.6		/ _[-voice]	<b>0.36</b>	<b>&lt;.05</b>

Table 3.11 - Correlation coefficients between the variables *duration* and *F1/F2* in the SFVs as obtained by a Spearman rank correlation (where rho = correlation coefficient). Correlations which reach statistical significance below the 5% level are given in bold.

It seems clear that there is nowhere a statistically significant correlation between F1 and vowel duration. However, some F2 effects can be noted: In the KIT data there is a positive correlation between F2 and duration in monosyllables and before voiced stops, which is in keeping with the above analysis. Within the DRESS/TRAP data F2 correlates with duration overall, although the effect seems to be carried by monosyllables only. In addition, there is an interesting mismatch between the two vowels in terms of how they behave across the two speaker groups: Whereas duration and F2 correlate in DRESS in the early group only, the inverse is the case in the TRAP data.

### 3.6.4 Discussion – Vowel duration

The important points seem to be:

- DRESS and TRAP are considerably longer than KIT.
- For all three lexical sets, vowels before voiced consonants are longer than before voiceless ones.
- In addition, this effect is stronger in DRESS than it is in any of the other two vowels.
- Lengthening factors combine more linearly in longer segments.
- There is no statistically significant effect of F1 (i.e. height) on vowel length.
- In monosyllables, there is a strong effect of F2 on vowel length, whereby, on the whole, frontness correlates positively with length.
- TRAP differs from the other two vowels in that the F2 effect is more important in predicting vowel duration than the voicing state of the following consonant.
- There are no statistically significant duration differences between the two groups in any of the SFVs.

From these results, a number of important implications follow with regard to front vowel length in the Intermediate Period. First of all, overall vowel length indicates that DRESS and TRAP are in fact different from KIT, since they are appreciably longer. Against the background of the front vowel shift, we can therefore conclude that they are indeed ‘different things’ on the grounds of their differential behaviour along two phonetic dimensions (i.e. DRESS and TRAP both raise/front *and* are longer; KIT centralises and is shorter).

Secondly, the lack of any strong correlation between F1 and length permits two possible interpretations of the relation between height and quantity in NZE. The straightforward one would be to conclude that in NZE, intrinsic constraints on vowel length depending on the vowels openness are simply overridden. The appeal of this assumption probably depends on how mechanistically one interprets such a constraint in the first place, that is, whether this relation is primarily language- or dialect-specific or whether it follows from motoric constraints in speech production. Under the latter premise, this interpretation of

vowel length in NZE becomes rather ad hoc and mysterious. The alternative would be to assume that the condition holds, but that there has been moderate subphonemic lengthening over time in the SFV set in NZE that offsets this intrinsic effect and thereby preserves isometry in the face of movement across articulatory space. However, the data presented here permit no verdict in this matter<sup>27</sup>.

We noted in section 3.5.4 that the allophonisation in the F2 dimension is rather more pronounced in KIT compared to the other two vowels. Against the background of the duration analysis, this seems plausible: Recall that Stevens and House (1963, cf. section 3.5.4 above) found that coarticulatory effects in the F2 dimension whereby fronting in velar environments as well as backing in fricative environments was strongest in short vowels. As for the velar environment, this effect is presumably due to transitional articulator movements taking up a larger portion of the overall duration of a vowel in short vowels compared to long vowels, assuming isochrony in the velocity of articulator movements. Extending this to the duration properties of the SFV in the present study, we would therefore expect the shortest vowel (KIT) to show the largest degree of allophonisation, which is the case. This is further evidence for the interrelatedness of the duration dimension and the F1/F2 dimension. In other words, it is not surprising that it is in the development of KIT where we find better explanations on coarticulatory grounds.

With respect to Labov's theory of directional vowel shift, the present analysis casts doubt upon the alleged exceptional status of DRESS and TRAP in NZE. Apart from the structural points briefly mentioned in the introductory section, we can also note that 'phonetic shortness' of DRESS and TRAP is hard to ascertain in the face of the presence of a phonetically much shorter vowel, namely KIT. Apart from phonetic duration, it was also shown that DRESS and TRAP show a number of phonetic interactions which are more typical of long vowels, such as a larger duration difference depending on whether the vowel occurs in mono- or polysyllables or before a voiced or voiceless stop. It may therefore be advisable to remain cautious in positing that DRESS and TRAP are short vowels in NZE (and, in fact, any standard variety of Modern English) and should therefore be expected to behave like LOT, STRUT, or FOOT rather than their tense counterparts. Although it is true that they originate as short

---

<sup>27</sup> It should be pointed out, however, that this hypothesis is not in principle untestable. One would need to show either one of two things: That before the Intermediate Period, lengthening and raising of DRESS and TRAP occurred at different times or in the speech of different individuals, or that in present-day NZE, lengthening and raising occur independently of each other.





## J.M. (Early male) KIT/DRESS - F1/duration

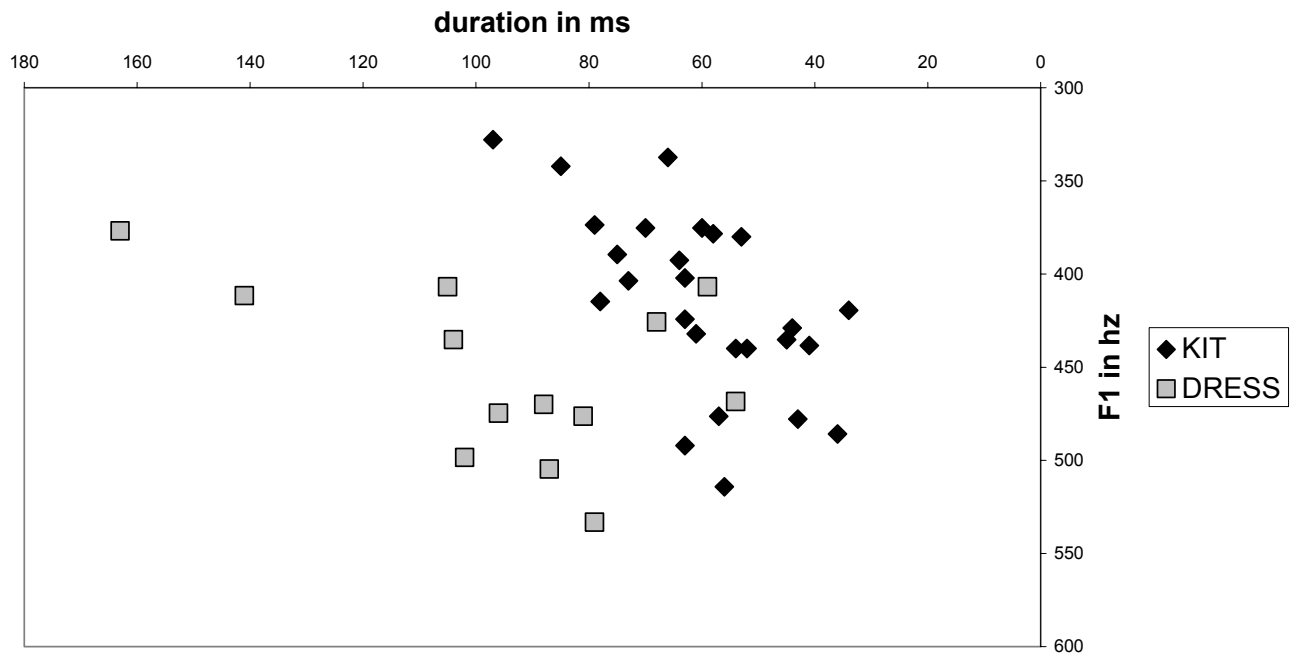


fig 3.17 (b)

Figure 3.17 – Differential degrees of overlap between the lexical sets of KIT and DRESS in the speech of an early male speaker. Figure 3.17 (a) plots the first two formants against each other, 3.17 (b) plots duration against the first formant. Note that this figure shows a restricted data pool from this speaker compared to figure 3.6.

Figure 3.17 (a) shows that for one and the same individual, a number of tokens are ambiguous between DRESS and KIT in F1/F2 space. However, if we plot length against the first formant frequency for the same speaker (fig. 3.17 (b)), this overlap largely disappears. It therefore appears that there is, to some extent, a trade-off relation between duration and formant frequency in terms of vowel overlap. Figure 3.18 below plots duration against both formant frequencies for the early male and the late female sample.

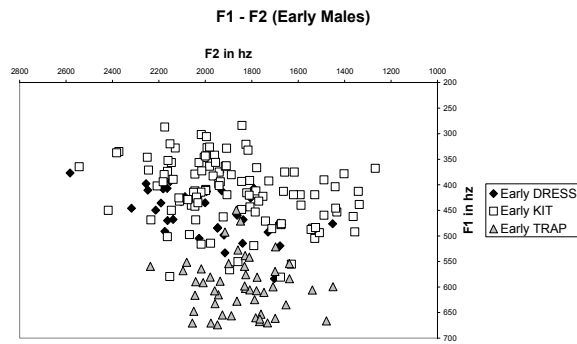


fig. 3.18 (a)

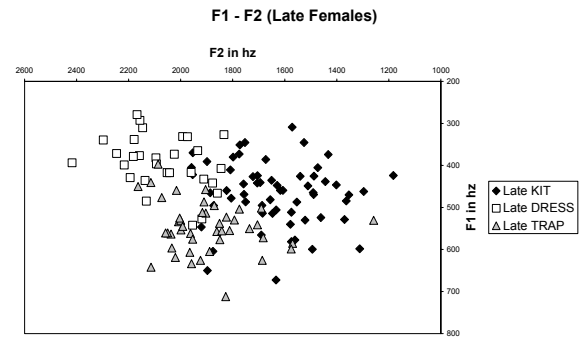


fig. 3.18 (d)

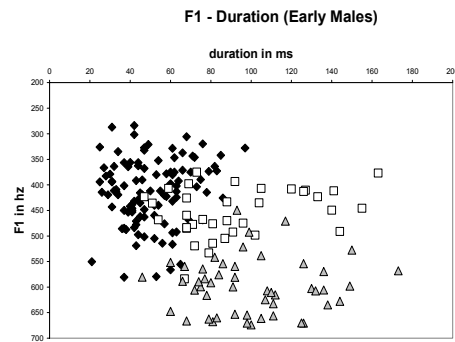


fig. 3.18 (b)

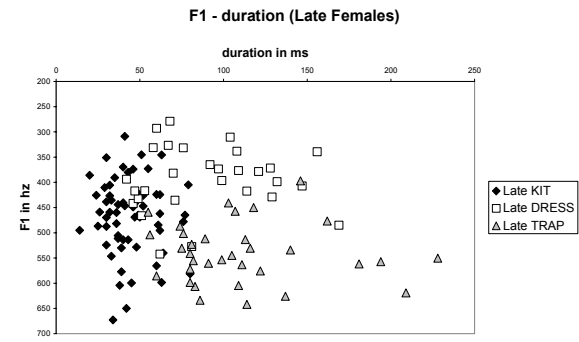


fig. 3.18 (e)

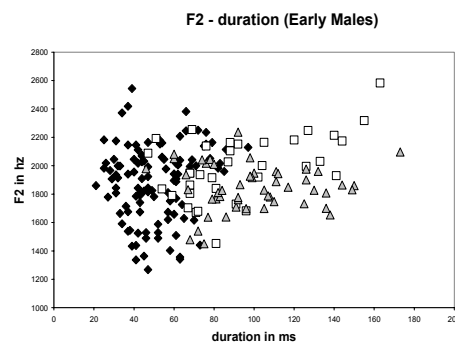


fig. 3.18 (c)

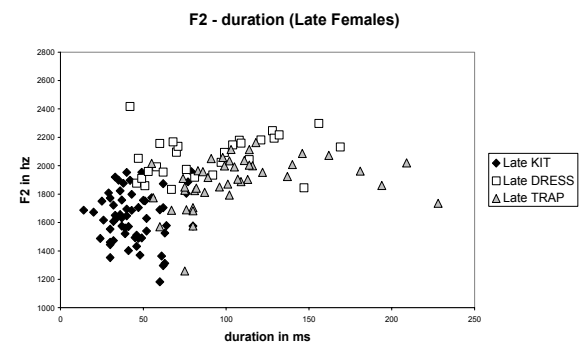


fig. 3.18 (f)

Figure 3.18 – F1/F2/duration plots of KIT, DRESS and TRAP in the early male and late female sample.

In the F1/F2 dimension, the early sample shows nearly complete overlap between DRESS and its neighbouring vowels. As was shown above, this overlap is resolved over the Intermediate period by KIT centralisation and DRESS raising (3.18 (b)). However, if we plot frequency against duration (figs 3.18 (c) – (f)) it appears that the TRAP distribution moves toward the KIT distribution in the F1/duration plot. Table 3.12 states Hotelling-Lawley trace scores for each of the distributions plotted in figure 3.18.

		<b>F1 – F2</b>	<b>F1 – duration</b>	<b>F2 - duration</b>
<i>Early Males</i>	KIT – DRESS	0.17	0.97	0.79
	DRESS – TRAP	1.99	2.07	0.2
	KIT - TRAP	2.11	3.58	1.33
<i>Late Females</i>	KIT – DRESS	1.53	1.07	1.99
	DRESS – TRAP	1.6	1.6	0.4
	KIT - TRAP	0.95	1.57	1.52

Table 3.12 – Hotelling-Lawley trace scores of the distributions of KIT, DRESS and TRAP in the early male and the late female sample. All scores are significant below .01.

The H-L traces shown in table 3.12 indicate wider separation between KIT and DRESS in the late sample compared to the early sample. As for DRESS vs. TRAP, it is only in the F1/duration dimension where the distance between the two distributions decreases over time. Finally, KIT and TRAP approximate each other in all dimensions.

The assumption that speakers/listeners of a language that is undergoing a vowel change whereby the distributions of two vowel phonemes approximate each other over time take advantage of (or ‘exaggerate’) alternative cues to the identity of a token in the overlap region seems plausible. What seems to be important here is that subphonemic durational cues can be dialect-specific and induced by historical processes such as vowel shift. In addition, the analysis presented above suggests that they are part of an ‘extrinsic allophony’ and are therefore variable in time and space.

What requires an explanation is why there should have occurred a reactionary movement of KIT, given that in this intermediate situation the two vowels were still solidly distinguished in F1 vs. duration space. The only hypothesis I have to offer here is purely conjectural: During the stage of the shift as exemplified by speakers such as J.M., we find three vowels in rather close proximity to each other, namely KIT, DRESS and FLEECE. What this amounts to is a three-way distinction in vowel-length in that region of the vowel space, which is not mirrored anywhere else in the vowel system. It has been pointed out before (cf. Bybee 2000 on the rise and fall of early Middle English rounded front-vowels) that although such constellations can arise as the outcome of non-teleological sound-change, they seem to be abandoned rather quickly. That is, phoneme systems tend to have similar types of

distinctions across the board. In the case of the SFV shift in the Intermediate Period, this means that the duration difference might be best regarded as a ‘transitional cue’ in disambiguating F1/F2 overlap between two phonemes.

### 3.7 Conclusion – The NZE front vowel shift

The preceding analysis of the SFVs has revealed the following picture: At the onset of the Intermediate period, we find a situation where DRESS is rather close to KIT, and that the latter vowel shows allophonisation in the F2 dimension where both fronted and centralised variants of KIT occur. This allophonisation is conditioned by phonemic environment, where the fronted variants are often adjacent to velar consonants, whereas centralised ones occur mainly in fricative environments. In addition, male speakers show a wider degree of allophonisation than do female speakers. This shows that although mean values indicate a pattern that is more similar to the initial state (cf. figure 3.1) in the speech of the early speakers, centralised variants were certainly already present at this stage. In addition, it was pointed out that there is considerable overlap in the F1/F2 dimension between KIT and DRESS, and much less overlap between DRESS and TRAP. Given this, the hypothesis that DRESS had undergone raising prior to the Intermediate Period makes sense and is in accordance with the conclusions in Gordon et al. (2004) and Trudgill (2004). In addition, we have observed that during the Intermediate Period DRESS and TRAP move away from each other, although both undergo raising (i.e. DRESS raises faster than TRAP). Plotting F1/F2 averages showed that while there are speakers with low DRESS and non-centralised KIT, the opposite case is not found, which is evidence for a *push-chain* that comes about *sequentially*. We also observed that the major predictors of formant frequency are phonemic environment as well as age, and that further factors such as lexical frequency, the existence of minimal pairs between SFVs, and syllable structure play only a minor role. Where they do show up in the regression analysis, their role is rather unclear.

It was shown in section 3.5 that the correlations between formant frequency and age suggest a somewhat finer picture of the overall changes in that, while DRESS raising and TRAP raising occur throughout the intermediate period, centralisation/lowering of KIT as well as

fronting of DRESS seem to occur consecutively. Section 3.6 gave a tentative account of the role of lexical frequency in the SFV shift, and concluded that innovativeness correlates negatively with frequency in the centralisation of KIT and the raising of DRESS, and positively with fronting of DRESS. In addition, it was shown that these effects are at best minor.

In section 3.7, the durational properties were analysed, and it was found that KIT is significantly shorter than the other two SFVs. In addition, it was pointed out that the highest degree of durational allophonisation is found in DRESS, where tokens in monosyllables before voiced stops are markedly longer than their complementary allophones.

Section 3.8 discussed the relationship between the formant frequency dimension and the duration dimension. It was found that those vowels that show the highest degree of overlap in the F1/F2 dimension are exactly those that show the most solid durational difference. This suggests that duration may have been a relevant dimension in the SFV shift.

## 4.

## The Front Centring Diphthongs - NEAR/SQUARE

## 4.1 Introduction

The front centring diphthongs /ɪə/ and /ɛə/ (NEAR and SQUARE) have been undergoing a phonemic merger in NZE. Although the merger is now very well documented (Batterham (1995), Gordon and MacLagan (1989/1996/1999/2001), Holmes and Bell (1992), MacLagan and Hay (2006)), there has been considerable debate regarding its historical origins with regard to (a) when did the two sounds started approximating each other, (b) what the initial direction of the merger was, (c) what type of merger it has been.

The present chapter will investigate these points on the basis of the data from the Intermediate Archive. Section 4.2 gives some background information regarding the historical origins of the front centring diphthongs (henceforth FCDs) as well as reviewing previous research on the merger of the FCDs in NZE. In addition, section 4.3 will outline some more general points regarding different types of mergers. Section 4.4 – 4.5 will present and discuss data on nucleus (point) measurements of the FCDs from the current sample. Section 4.6 will be concerned with the dynamic properties of NEAR and SQUARE. Differences in the realisation of the front centring diphthongs with respect to different degrees of stress are discussed in 4.7. Section 4.8 will focus on the distributional properties of the data and shed light on the question of what type of merger the NEAR/SQUARE merger is an example of. A brief discussion of lexical frequency effects is presented in section 4.9. Section 4.10 analyses durational properties of NEAR and SQUARE.

## 4.2 Background

*Historical origins*

The centring diphthongs (in what follows, this term refers to NEAR and SQUARE, i.e. those with a front nucleus) are a relatively recent addition to the phoneme system of English. Their

occurrence is furthermore restricted to the so-called ‘non-rhotic’ varieties of English, i.e. varieties that have lost non-prevocalic R (which are, in rough geographical terms: British English outside Scotland, Ireland and the South-West of England; the Southern Hemisphere Englishes<sup>28</sup>; Southern and New England varieties of American English). Before the onset of loss of non prevocalic R, they were pre-rhotic allophones of ME/EModE long front vowels. The relations between the centring diphthongs and their historical allophones, however, is not quite straightforward since mergers affected each of the two subsets in a somewhat different fashion. In brief, NEAR derives from ME pre-rhotic /e:/ or /ɛ:/, while SQUARE derives from ME pre-rhotic /e:/, /ɛ:/, /a:/ or /ai/. Spelling is suggestive here: -ere- and -eer- spellings (cf. ‘here’, ‘there’, ‘beer’) usually indicate ME /e:/, -ear- indicates /ɛ:/ (‘fear’, ‘pear’, ‘spear’), -air- indicates ME /ai/ (cf. ‘air’, ‘fair’), and -are- indicates ME /a:/ (cf. ‘fare’, ‘care’). The predictability of their contemporary reflexes is limited due to a number of lexical idiosyncrasies, however, the following generalisations hold: ME pre-rhotic /e:/ comes down as NEAR (exceptions: ‘there’ and ‘where’), /a:/ and /ai/ comes down as SQUARE across the board. /ɛ:/ is mostly NEAR, with the exceptions of: *tear* (V), *bear* (V+N), *wear*, *swear*, *pear* (for more comprehensive accounts of vowel changes in Early Modern English, see Dobson (1957), Köckeritz (1953), Jespersen (1909), Luick (1914), Lass (1992)).

Recall from the introductory section of chapter 3 above that the current front vowel system of Standard English comes about as the result of a two-stage process which (a) raised the front monophthongs by one slot and subsequently (b) merged ME /e:/ and /ɛ:/ on /i:/, and /a:/ and /ai/ on /e:/. At some stage during or after these developments, the diphthongs must have become disjoint from their former non-pre-rhotic allophones. This is evidenced mainly by the quality of SQUARE in most Standard varieties of English: Whereas ME /a:/ and /ai/ have developed into upgliding diphthongs /eɪ/, this is not the case in their pre-rhotic allophones, i.e. their quality is now usually [eə] ~ [ɛə]. At this stage, the question arises as to what the correspondence between the FCD nuclei and the monophthong system is. It is usually recognised that these nuclei are now equivalent to the SFVs KIT and DRESS. This, however, can be doubted on the basis of the recent developments in NZE. A comprehensive discussion of these issues will be given in section 4.14 - 4.16 of this chapter.

---

<sup>28</sup> This is an overidealised outline in both historical as well as synchronic terms: As has been documented before (Gordon et al. (2004)), rhoticity was rather widespread in the speech of 1<sup>st</sup> and 2<sup>nd</sup> generation speakers of NZE. The same holds in earlier Australian English (Trudgill, p.c.). In addition, rhoticity is still present in the South of New Zealand, in somewhat restricted phonological environment (Bartlett, p.c.).

*The FCDs in related varieties*

As phonemes, FCDs exist only in non-rhotic varieties of English, i.e. varieties which have lost non-prevocalic /r/. Although schwa epenthesis occurred independently of /r/ loss (Wells 1982: 153), the phonemic status is hard to ascertain in rhotic varieties due to the paucity of potential minimal pairs (*ibid.*: 154). As far as the opposition between NEAR and SQUARE is concerned, mergers between the two phonemes have also been reported from East Anglia, the US South, and the West Indies (Wells 1982:157). In addition, monophthongal variants of both diphthongs (but no merger) occur in Australia. We therefore assume a merger of NEAR/SQUARE that arose independently in NZE.

*The FCDs in contemporary NZE*

NZE has undergone a phonemic merger of the FCDs, the major patterns of which are straightforward, whereas the details are rather complex. The phenomenon is well-studied by now, and the current consensus seems to be that for people who have the merger in production, the merger is on the higher vowel (Gordon and MacLagan (1985/1989/2001)). Earlier reports (Holmes and Bell (1992)) found merger on the lower vowel. It has also been pointed out that the merger must have been a rather recent development (i.e. after 1900), since MU speakers keep the two phonemes apart (cf. Gordon et al. (2004), Watson et al. (2000)). In addition, there is now growing evidence that the process may be, in part at least, a so-called ‘near-merger’ (cf. Labov 1994), whereby there are speakers who make the difference in perception, but not in production (for perceptual data cf. Warren et al. (2004), Warren (2006)).

What has received less attention are the mechanics of the change. More specifically, it is unclear whether the NEAR-SQUARE merger is a merger of approximation, or whether it proceeds via transfer of lexical items (more on this below). In addition, comments regarding the phonetic conditioning are few and far between (e.g. MacLagan and Gordon (1996) claim



that in its initial stages, the merger was most advanced in environments of low stress; Warren (2006) states that the merger affected vowels following coronal consonants first.<sup>29</sup>).

### 4.3 Mechanisms of phonemic mergers

A phonemic merger between two phonemes A and B can be schematised as follows:

stage 1	stage 2
A	A/B
B	

In brief, two phonemes that are different at stage 1 (pre-merger) are no longer different at stage 2 (post-merger). This, however, leaves a whole range of questions open:

*What exactly is the status of the merged phoneme in relation to its predecessors?*

This relates to the directionality of the merger. By definition, different vowels occupy different places in a vowel space. Therefore, the element that is the outcome of a merger must be different from either A or B, or both. With regard to the schematic representation above, the question is whether we should designate the merged phoneme as (1) A, (2) B, (3) A/B, or (4) C. Most historical accounts of vowel mergers usually opt for either (1) or (2), which is to say that a merged phoneme is the result of some formerly distinct phoneme ‘becoming like’ the other phoneme. (4) is common in cases where a merger affects only a restricted number of

---

<sup>29</sup> Warren’s use of the term *coronal* is sufficiently idiosyncratic to require a brief comment: Warren states that ‘it was noted that the mid-age group had a smaller EAR/AIR distance for words beginning in postalveolar /ʃ/ and /tʃ/ and velar /k/, and that [...] since the velar is likely to be fronted before these front vowels, all three consonants can be characterised as coronal. The /eə/ vowel is also higher after /sp/, which involves the coronal /s/. It is questionable whether postalveolars should indeed be characterised as coronal. In addition, I am not aware of ‘fronted velars’ (i.e. dorso-palatals) having ever been claimed to be classified as [+coronal]. Furthermore, Warren seems to be positing transgressive post-coronal conditioning after /sp/, unless he regards /sp/ as a phoneme of English. This claim, however, is not independently verifiable since Warren’s study does not include the pairs *steer/stare*, *sneer/snare*. Finally, the pair *really/rarely* shows the largest distance between the nuclei of its members, although this is the clearest example of a [+coronal] \_\_ environment in Warren’s study.

allophones of two or more vowels, and where it can be shown that these merged allophones behave differently from their non-merged complements in the subsequent history of the language. An example would be the merger of ME short vowels /i/, /e/ and /u/ on /ɜ/ in pre-rhotic environments (the contemporary NURSE vowel).

*How does the merger come about?*

As with vowel shifts, mergers are processes that span over a given period of time. Therefore, you would expect in-between stages that may display properties that are different from both the initial stage and the final stage. Three different mechanisms are commonly recognised: (1) merger by approximation, which means that one of the formerly distinct vowel phonemes moves toward the space occupied by the phoneme it eventually merges with. This type of process is generally assumed to be both *gradual* as well as *phoneme-based*, i.e. throughout the entire shift, the vowel space occupied by each of the two phonemes remains approximately constant, whereas that of the two phonemes together shrinks. (2) Merger by transfer, whereby the vowel space occupied by each of the two phonemes does not change until the transfer is complete. Rather, lexical items change their vowel phoneme abruptly until all words that used to contain a given vowel now contain the vowel it merges with. (3) merger by extension. This refers to an extension of the vowel space occupied by each of the two phonemes until they overlap.

The details of these different mechanisms have been opened up to empirical investigation only recently, since they require direct evidence from a data pool where the merger is actually underway. Historical reconstruction based on internal reconstruction or cognate matching cannot, in principle, falsify a given proposed mechanism in a given instance of a merger. Note that (2) may also arise artefactually as an outcome of (1): If a given vowel distribution V1 moves in the direction of another vowel distribution V2, and assuming that V1 is to some extent allophonically structured, certain allophones will reach the accretion zone of V1 earlier than others, in which case there may obtain an intermediate situation where words containing allophones that are nearer to the V2 distribution would contain the new vowel, whereas those containing allophones that are further from V2 would not, which might be an instance of ‘illusionary’ lexical transfer. As I will show below, these problems are empirically assessable, and given an appropriate data pool, can be resolved.

A phenomenon that has received growing attention in recent years is that of *near-merger*. This refers to either one of two related phenomena: (1) Two vowels merge in perception, but not in production; (2) two vowels overlap *nearly* completely in vowel space, but the small differences suffice to maintain their etymological independence. The finding that near-mergers occur has been hypothesised to explain a number of historical phenomena which remained puzzling under previous accounts. A widely recognised example includes the non-merger of ME /i:/ and /ai/ after the GVS. One outcome of the GVS was that ME /i:/ diphthongised to /ɪi/ and subsequently moved to /ai/. At the same time, however, ME /ai/ rose to /eɪ/, which implies that there should have occurred a cross-over of the nuclei at some stage in the process, and hence, the two sounds should have merged, which is not the case (i.e. FACE and PRICE are different). The account advocated by Labov (1994) posits that this behaviour might be explained by a near-merger of the two phonemes<sup>30</sup>.

#### 4.4 NEAR / SQUARE in the Intermediate period – Static properties

In this section we will investigate the properties of the FCDs in F1/F2 space. We will focus on the results obtained from point measurements of the nucleus. 4.4.1 will show some general patterns, 4.4.2 provides a regression analysis of a number of linguistic factors as well as speaker sex and age.

##### 4.4.1 General patterns

Figure 4.1 below plots the mean F1/F2 of NEAR and SQUARE of all Intermediate speakers relative to the means of FLEECE, KIT and DRESS. Figure 4.2 (a) – (f) plots each of the six subsamples.

---

<sup>30</sup> This, however, is not the only account. Lass (1990) states that ME /ai/ monophthongised to /e:/ before rebreaking to /eɪ/, which makes sense in light of the merger of this phoneme with ME /a:/, which was originally a monophthong.

## Intermediate Front Vowels

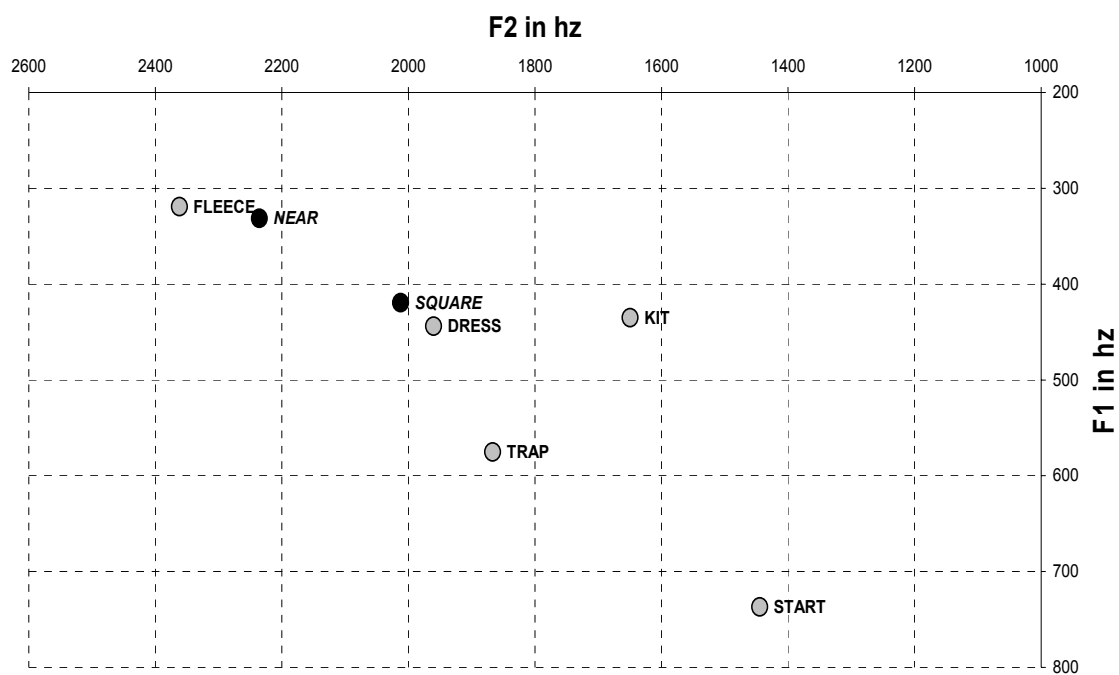


Figure 4.1 – Front vowel formant frequency averages in the Intermediate sample.

## Early Males - Front Vowels

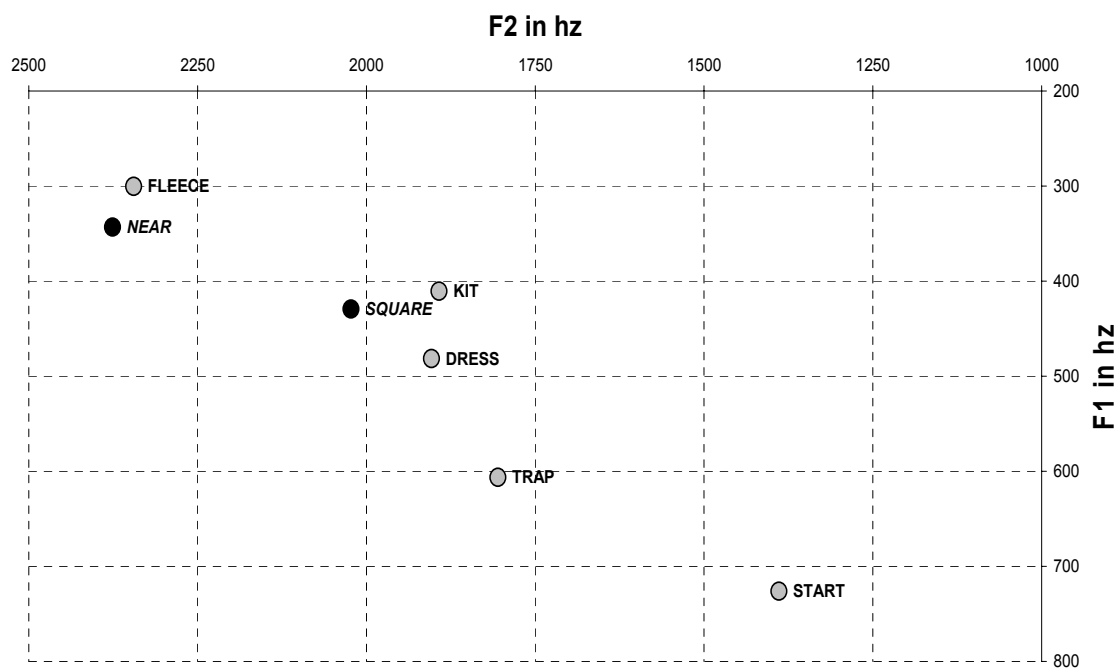


Figure 4.2 (a)

### Early Females - Front Vowels

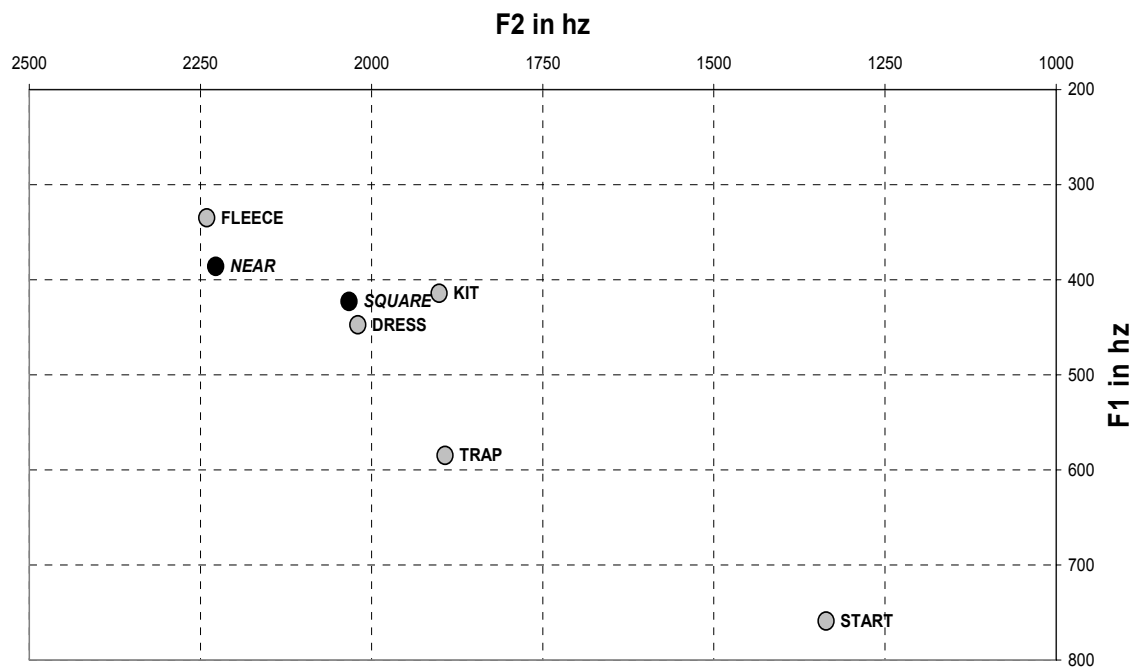


Figure 4.2 (b)

### Medium Males - Front Vowels

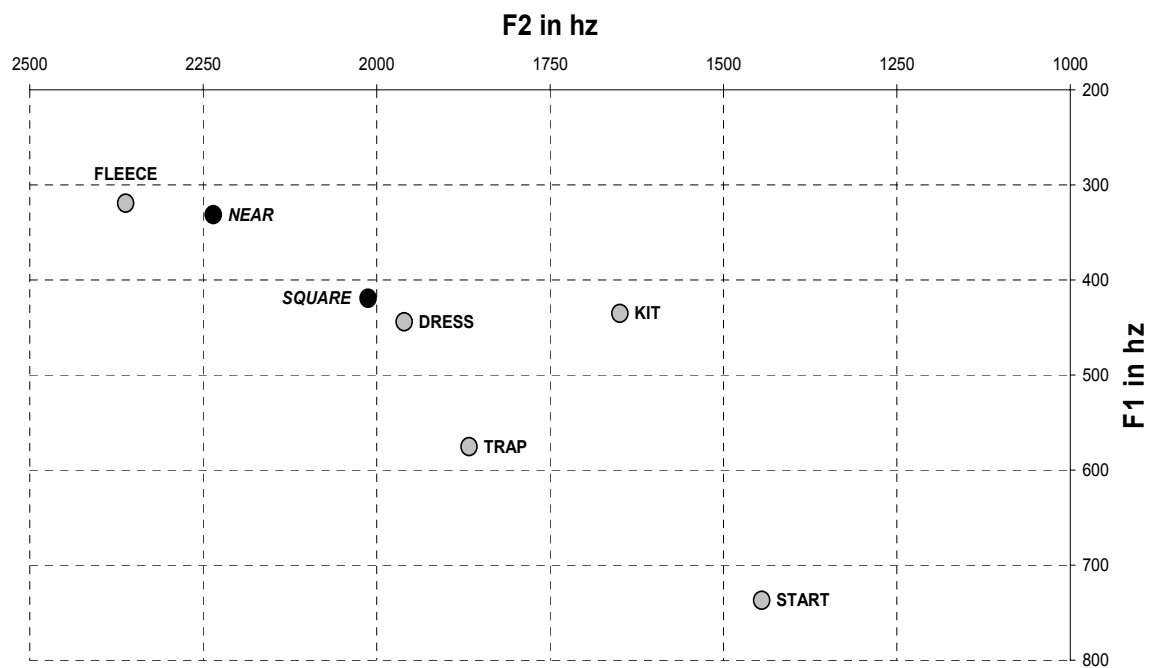


Figure 4.2 (c)

### Medium Females - Front Vowels

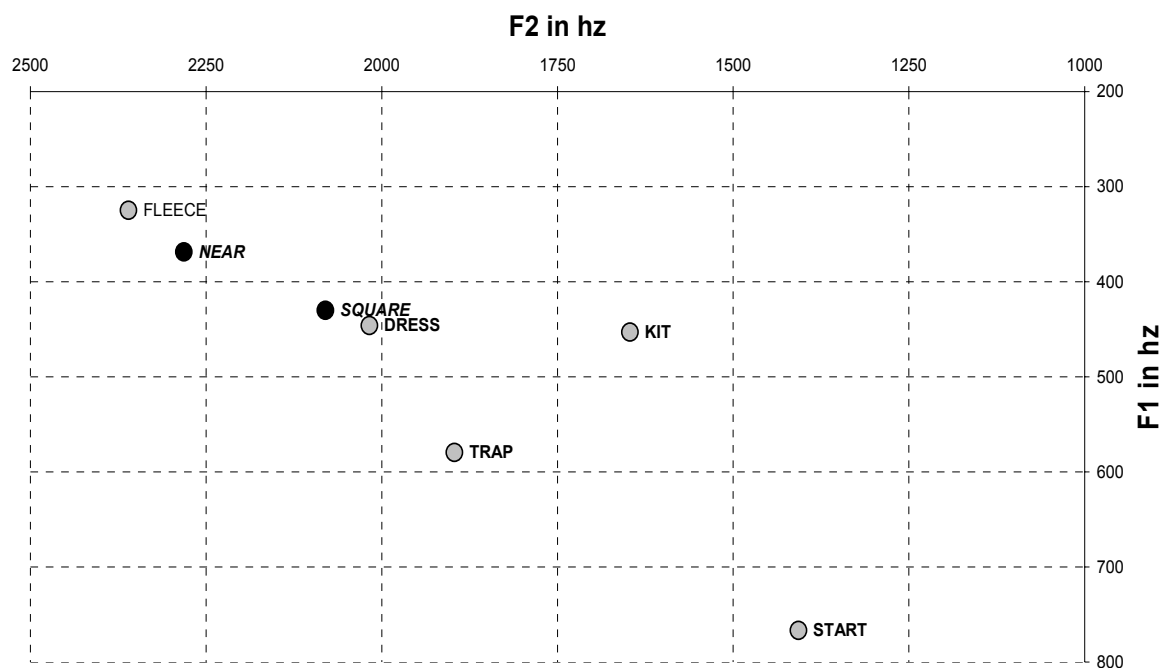


Figure 4.2 (d)

### Late Males - Front Vowels

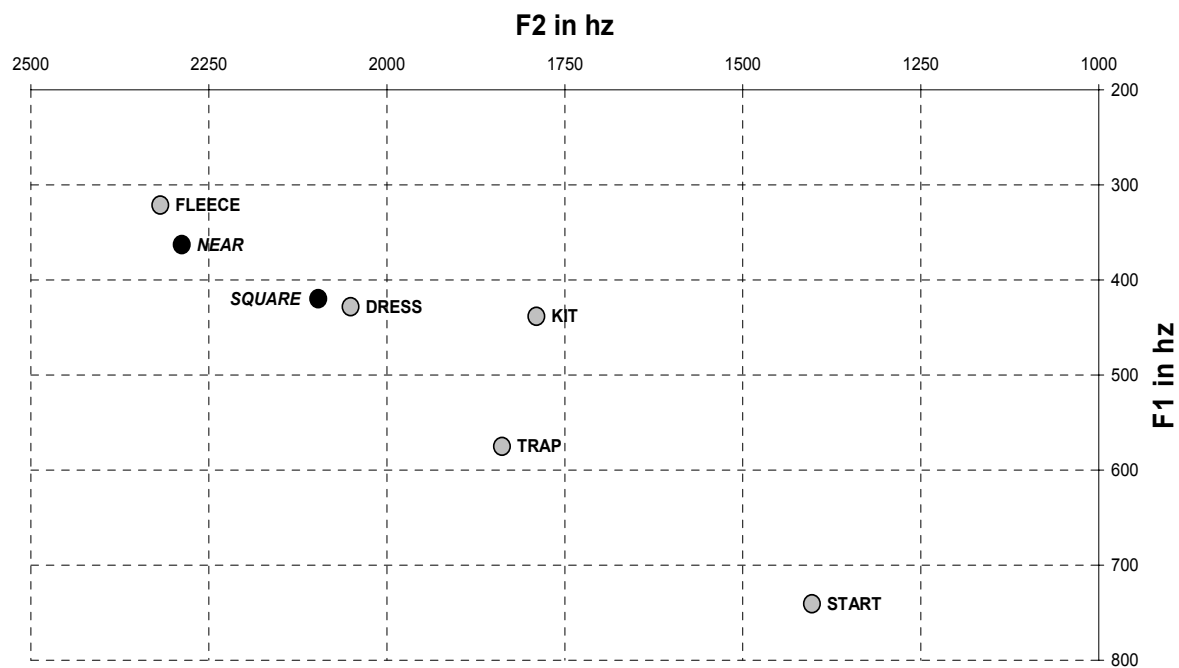


Figure 4.2 (e)

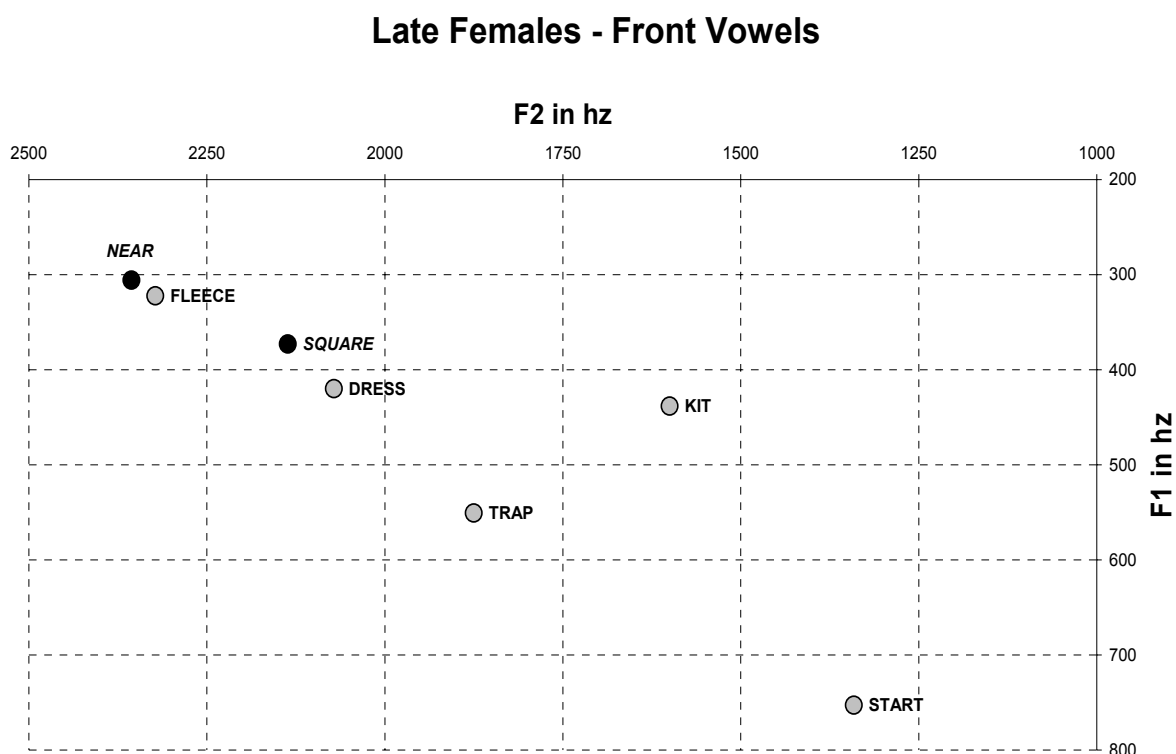


Figure 4.2 (f)

Figure 4.2 – Front vowel formant frequency averages in the Intermediate sample, broken down into age and speaker sex: Early males (4.2 (a)), early females (4.2 (b)), medium males (4.2 (c)), medium females (4.2 (d)), late males (4.2 (e)), late females (4.2 (f)).

Figure 4.1 shows that the position of the nucleus of NEAR is comparatively close to that of the long high monophthong FLEECE, while the closest monophthongal correlate of SQUARE is DRESS. In addition, the two FCDs seem to be kept distinct in the Intermediate period. Interestingly, however, the position of SQUARE relative to DRESS indicates that, at any stage in the Intermediate period, the relation between the two vowels is different to what we would expect on etymological grounds. As was pointed out in section 2 above, SQUARE is the reflex of pre-rhotic allophones of ME long low-mid and low vowels. It has been pointed out that at some stage after the GVS, the nucleus of SQUARE became associated with the DRESS vowel (this is implied by transcriptions such as / $\epsilon$ / for DRESS and / $\epsilon\text{ə}$ / for SQUARE), while many pre-rhotic allophones of DRESS split off and have fallen into the lexical set of NURSE, mostly in monosyllabic environments. The data from the Intermediate period supports this hypothesis

well, since we have raised realisations of both DRESS and SQUARE. Furthermore, this parallelism is not reflected in NEAR vis-à-vis KIT. This is to some extent unexpected, since it is generally assumed that the nucleus of NEAR relates to KIT as that of SQUARE does to DRESS. Hence, the absence of centralisation in NEAR remains to be accounted for. This problem will be discussed at length in chapter 5 below.

If the data is broken down into the six subsamples (figure 4.2), it appears that within the male group there is a clear pattern of decreasing distance between NEAR and SQUARE over the Intermediate period. The Euclidean distances between NEAR and SQUARE are as follows: 363.7 hz (Early Males), 239.9 hz (Medium Males), 199.9 hz (Late Males). This suggests that there certainly seems to be a pattern of approximation between the centring diphthongs over the Intermediate period. However, the exact relationship between the two vowels is less clear. It appears that between the early male and the medium male sample, the decreasing distance is mainly due to the movement of NEAR to a more central position, whereas the late male sample shows fronting of SQUARE. In addition, the position of NEAR relative to FLEECE and KIT does not show a clear directional pattern over time (in the early and late male samples, NEAR is lower than FLEECE, whereas it is more central in the medium sample). The position of SQUARE remains rather more stable in relation to DRESS, where the former is slightly more fronted/raised.

This situation is furthermore complicated if we look at the female sample: Although the increments are rather small, it seems as if there was an actual increase in distance between NEAR and SQUARE over time in the female sample (Euclidean distances are 198.1 hz (early females), 210.4 hz (medium females), 229.5 hz (late females), respectively). These numbers are in striking disagreement with both expectations one would have regarding the directionality of a sound change of this type as well as the auditory assessment of the measured tokens: In general, it is in the later samples where a number of SQUARE tokens were auditorily coded as NEAR, and, although more rarely, vice versa. However, we observed in chapter 3 that distance measures between mean positions of vowels may not reveal a number of pertinent patterns in their development over time, and that distributional properties should be taken into consideration as well. After all, distances between means do not reveal degrees of allophonisation, nor overall distributional shapes. Table 4.1 shows Hotelling-Lawley trace scores for each of the six subsamples:



<i>speaker group</i>	<i>hotelling-lawley</i>	<i>p</i>
<b>Early females</b>	0.213	< .0001
<b>Medium females</b>	0.454	< .0001
<b>Late females</b>	0.1114	< .05
<b>Early males</b>	1.308	< .0001
<b>Medium males</b>	0.605	< .0001
<b>Late male</b>	0.3136	< .0001

Table 4.1 – Hotelling-Lawley traces and corresponding significance levels for NEAR - SQUARE distributions in six groups of Intermediate speakers.

With respect to the male speakers, the pattern we observed in figure 4.2 is reflected in table 4.1 in that there is decreasing distance between the distributions of NEAR and SQUARE over the Intermediate period. In addition, the distributions remain different at p-values below the 0.01% level of significance. While the pattern of increasing distance between the early and the medium female sample is also found in table 4.1, the relation between these and the late female sample is intriguing: In table 4.1, it appears as if the late female sample shows both the smallest distance between the distributions of NEAR and SQUARE as well as a markedly lower p-value, i.e. the difference between the two distributions is less significant. This is in keeping with both the auditory assessment of NEAR/SQUARE across the subsamples as well as what we would expect on the grounds of findings in chapter 3 (where the late female group showed the most innovative patterning of KIT, DRESS and TRAP) as well as general sociolinguistic considerations (since females generally show more innovative realisations of a non-stigmatised sociolinguistic variable).

It therefore seems as if the precursors of the merger between NEAR and SQUARE as evidenced by the data from the Intermediate sample are a complex interrelation of different developments, namely (1) fronting/raising of SQUARE, (2) an expansion of the vowel space occupied by both centring diphthongs and (3) a less directional development of NEAR. That

(2) is the case is shown in table 2 below, which shows standard deviations of both centring diphthongs relative to their mean in all six speaker groups.

	<i>NEAR F1</i>		<i>NEAR F2</i>		<i>SQUARE F1</i>		<i>SQUARE F2</i>	
	<b>avg</b>	<b>SD</b>	<b>avg</b>	<b>SD</b>	<b>avg</b>	<b>SD</b>	<b>avg</b>	<b>SD</b>
Early males	330.7 hz	70.7 hz	2394 hz	182.9 hz	432.3 hz	52.2 hz	2005.8hz	206.6 hz
Medium Males	335.9 hz	46.3hz	2208.6hz	425.8 hz	419.6 hz	60.8 hz	2020.4hz	319.2 hz
Late males	363.5 hz	118.9 hz	2260.2hz	210.3 hz	410.8 hz	74.6 hz	2078.8hz	163.3 hz
Early females	374.8 hz	61.1 hz	2209.8hz	281.5 hz	417 hz	61.4 hz	2028.8hz	240.3 hz
Medium females	364.5 hz	61 hz	2300.9hz	171.9 hz	430.6 hz	53.8 hz	2088.9hz	247 hz
Late females	329.7 hz	98.5 hz	2326hz	186.5 hz	372.3 hz	65.2 hz	2174.9hz	371.5 hz

Table 4.2 – Average formant frequencies and standard deviations of the nuclei of NEAR and SQUARE in each of the six groups of Intermediate speakers.

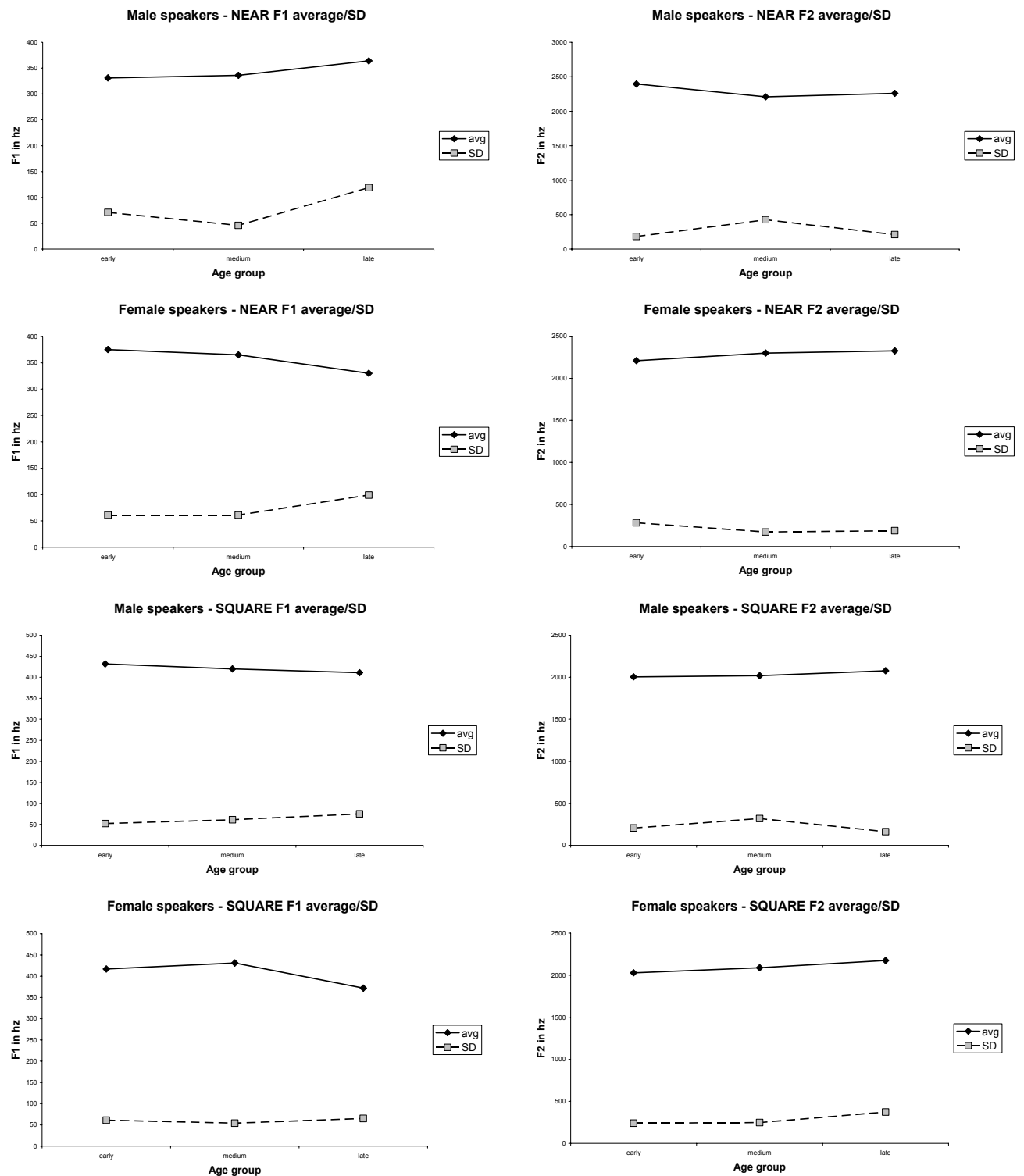


Figure 4.3 – Mean formant frequency and SD in six groups of Intermediate speakers.

Figure 4.3 represents the data from table 4.2 as age/formant frequency plots. It appears that in the F1 dimension of NEAR, there is a clear trend toward higher SDs in the late samples, while the development in F1 is different for male and female speakers in that the male speaker show lowering of F1 over time, whereas the opposite situation holds in the female sample. In both cases, however, a comparison between early and late speakers suggests a less well-specified nucleus of NEAR (i.e. for the males, NEAR becomes further spread out in articulatory space, whereas for the female speakers, it approaches SQUARE in the F1 dimension).

As for SQUARE, there is a trend toward lower F1 values in both the male and the female samples, with a concomitant increase in SD in the male sample. The F2 data is somewhat more confusing: Whereas there is a trend toward lower F2 values for NEAR in the male sample, the opposite holds true in the female sample as well as for SQUARE in both samples. Little directionality is found in the development of the SDs over time in the F2 dimension.

This data therefore suggests that the loss of distinctness between NEAR and SQUARE in the Intermediate period is primarily a function of SQUARE raising and less well-specified distributions of both centring diphthongs in the F1 dimension. It also seems that to some extent, NEAR initially reacts to SQUARE raising in a chain-shift-wise fashion, i.e. by moving to a higher (male and female speakers) and more fronted (female speakers only) position.

#### 4.4.2 Factor analysis

The following section presents the results of a regression analysis in order to show which variables predict the Fn distribution of NEAR and SQUARE best, and in order to either corroborate or to falsify earlier claims as to the major phonetic influences on the initial stages of the centring diphthongs, especially Warren's claim that the merger started in coronal environment, and Gordon's claim that the merger originated in tokens of low stress. Figure 4.4 (a) represents pooled F1 data, 4.4 (b) shows F2. The independent variables are: Age - speaker sex - vowel - stress - preceding/following phoneme - place/manner/voicing of articulation of the preceding/following phoneme - major place feature of the preceding/following phoneme.

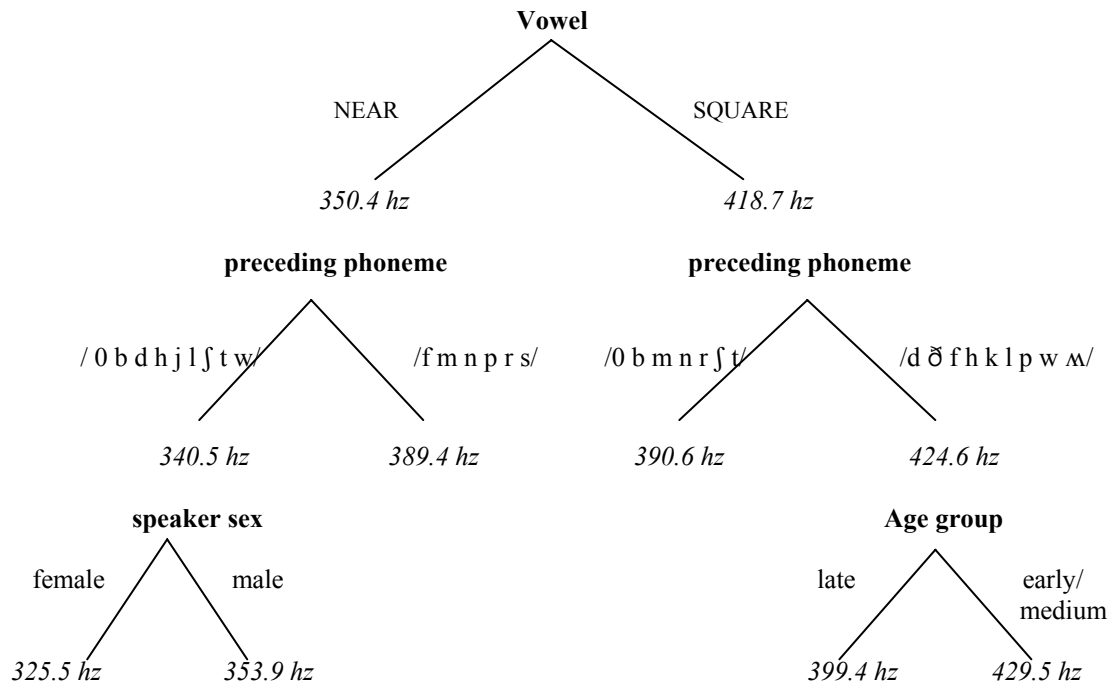


Figure 4.4 (a)

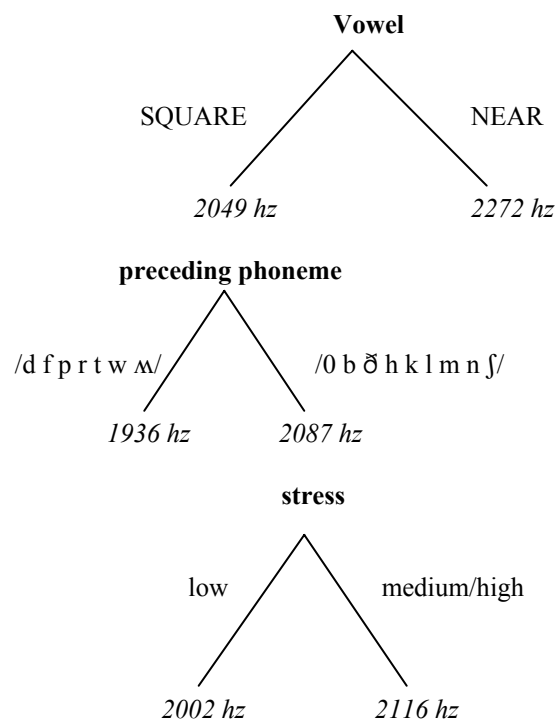


Figure 4.4 (b)

Figure 4.4 – CART representations of major predictors of pooled NEAR/SQUARE data in the Intermediate sample. 4.4 (a) shows F1 data, 4.4 (b) shows F2.

In both the F1 and the F2 data, the major split is in the variable vowel, i.e. NEAR and SQUARE are distinct in both height and backness. Within the F1 data, further splits occur along the variable preceding phoneme. The exact nature of this split is not straightforward since no higher-order generalisations regarding manner or place of articulation seem to play a role in partitioning the F1 data. The claim that vowels in coronal environments spearhead the change is only partially borne out: 4 out of 7 environments are [+coronal] in the innovative branch (i.e. the one with lower F1 values), while 3 out of 9 are [+coronal] in the more conservative branch (higher F1 values). Although the ratio of coronals is higher in the former branch, this effect is not strong enough to override the effect of individual phonemes in predicting F1 for SQUARE. As for NEAR, the ratios are 4/9 in the branch with low F1 values, and 3/6 in the complementary branch. While this suggests that the distribution of vowels in [+coronal] and [-coronal] environments is almost random, it is furthermore not clear what should count as ‘innovative’ within NEAR. On the one hand, it was suggested above that for some speakers at least (i.e. the female speakers), NEAR enters into a transitional chain-shift relation with SQUARE in the F1 dimension. From this follows that for these speakers, NEAR tokens with lower F1 values can reasonably be regarded as more innovative if we restrict the perspective to the intermediate (female) sample only. However, a post-hoc perspective onto the overall process to date suggests that innovativeness in NEAR is probably best equated to similarity to SQUARE (which also seems to be implied by the distance measure in Warren (2006)), i.e. tokens with higher F1 values. We therefore conclude that the special role of vowels in coronal environments is not a strong factor in the data from the Intermediate period.

Within the F1 data we also find non-linguistic effects within both the NEAR branch and the SQUARE branch. More specifically, female speakers show closer realisations of NEAR than males. This effect, however, is restricted to a number of allophones. Within the more conservative branch of SQUARE (i.e. those allophones with a higher average F1) we find partitioning along the variable age, whereby late speakers show lower F1 values than earlier speakers.

The analysis of the F2 data (figure 4.4 (b)) reveals a similar structuring: The major branch splits off NEAR and SQUARE from each other. Furthermore, the SQUARE data is partitioned along the variable preceding phoneme, where vowels after /d f p r t w m/ show

more central realisations than vowels after /θ b ð h k l m n ʃ/. As with the F1 data above, no special role can be attributed to vowels in coronal environments. We finally note a stress effect within the latter branch whereby tokens of low stress show more central realisations than those with medium and high stress. Since frontness in SQUARE corresponds to similarity with NEAR, the claim that the merger is more advanced in low-stress environments is not supported in this analysis. However, since the stress effect is restricted to a range of allophones in the F1 dimension of SQUARE only, a more thorough analysis is required, which will be carried out in section 4.9 below.

The pooled data suggests that in the Intermediate period as a whole, the major predictor of formant frequency is vowel identity, which shows that the two vowels are still significantly different in this period. However, I suggested above that there are reasons to assume that there is evidence for at least an approximation of the nuclei of the two centring diphthongs. It is therefore reasonable to ask to what extent the analysis on pooled F1/F2 data is reflected if we restrict the data range to smaller subsamples. The analysis of distance measures between distributions (cf. table 4.1 above) suggested that the early males show the largest distance between NEAR and SQUARE, while the late females show the smallest distance. If it is the case that the late female speakers show precursors of the merger, we would probably expect this group to show a partitioning of formant frequency data where vowel identity plays less of a role than is the case in the most conservative group. Figure 4.5 plots CART representations of F1/F2 data for the early male and the late female group.

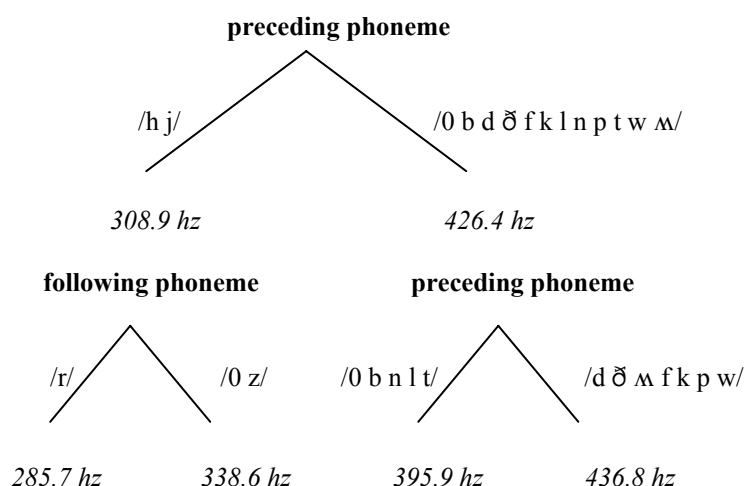


Figure 4.5 (a)

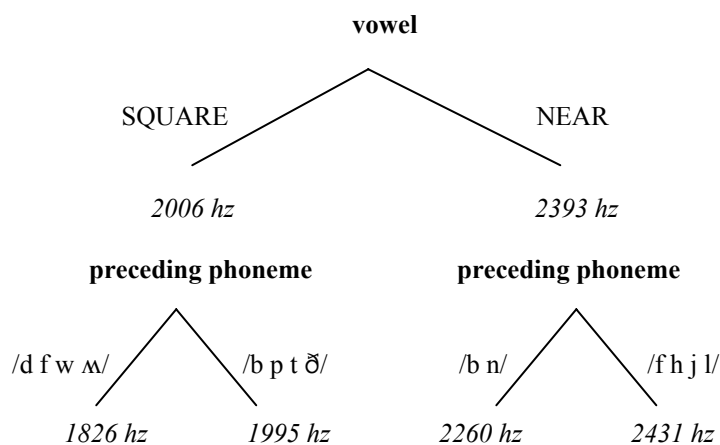


figure 4.5 (b)

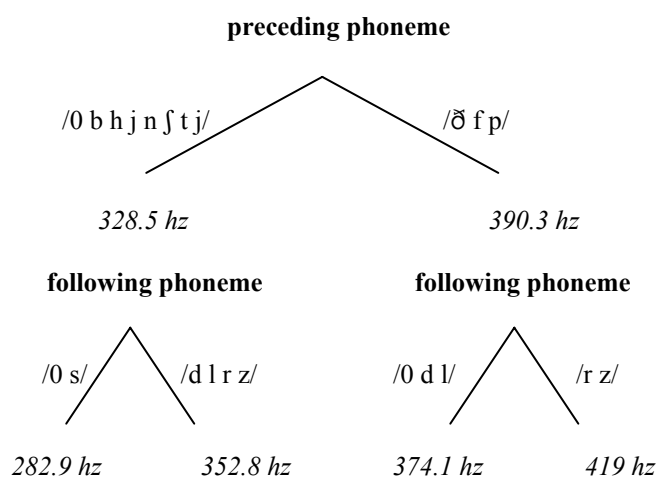


Figure 4.5 (c)

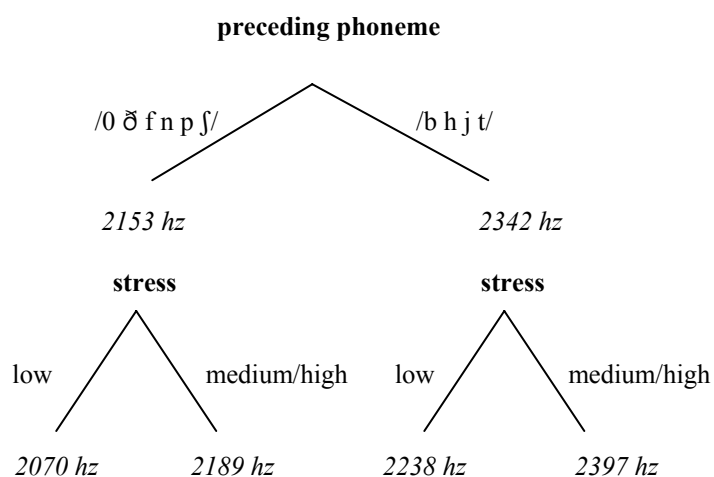


Figure 4.5 (d)

Figure 4.5 – CART representations of formant frequency data from the early male and the late female group. Figure 4.5 (a) shows the F1 data from the early male sample, 4.5 (b) F2 data from the same group. Figs. 4.5 (c) and (d) show F1 (4.5 (c)) and F2 (4.5 (d)) for the late female sample.



Interestingly, neither the early males nor the late females show a vowel identity effect in the distribution of the F1 data. This suggests that even at the early stages of the Intermediate period, the nuclei of the centring diphthongs are less solidly distinguished by height than was the case in, for example, DRESS and TRAP. In F2 (figs. 4.5 (b) and (d)) we note a vowel identity effect only in the early male sample, which is indicative of approximation of the FCDs in the front/back dimension over the Intermediate period.

#### 4.5 Discussion – Nucleus positions

Sections 4.4 analysed the position of the nuclei in F1/F2 space for each of the six subsamples as well as the major linguistic factors that predict F1/F2 distributions. It was shown that in each of the samples, the nucleus of SQUARE occupies a position which is rather close to DRESS; more specifically, slightly more fronted and raised than DRESS. The nucleus of NEAR behaves in a somewhat less predictable fashion. Although it seems clear that its closest equivalent within the monophthong system is FLEECE, the exact spatial relationships between NEAR and FLEECE are unclear. From this we infer that the assumption that the nucleus of SQUARE may have risen along with DRESS in NZE as a function of their equivalence seems well supported by the data from the Intermediate archive (but see chapter 5). The status of NEAR with respect to the monophthong system remains indeterminate.

The relation between the front centring diphthongs over the Intermediate period is one of decreasing distance as a partial function of both spatial approximation between vowels and spatial expansion within each of the vowels, i.e. this process is best reflected by distribution distance measures rather than distances of means. This suggests that we are faced with a pre-merger state in the Intermediate period, where the distributions of the two centring diphthongs are still significantly different, while the extent of that difference is comparatively minor, and other factors are better predictors of pooled formant frequency data.

## 4.6 Dynamic properties

### *Trajectories*

Having discussed the development of the nuclei of the front centring diphthongs in the Intermediate Period of New Zealand English, we will now turn to the results of the trajectory analysis. Figure 4.6 models the vowel trajectories of both vowels for all six speaker groups. The models were obtained by fitting lowess lines through time-normalised representations of the vowels. That is, each observation is plotted in its relative position within the diphthong token it occurs in.

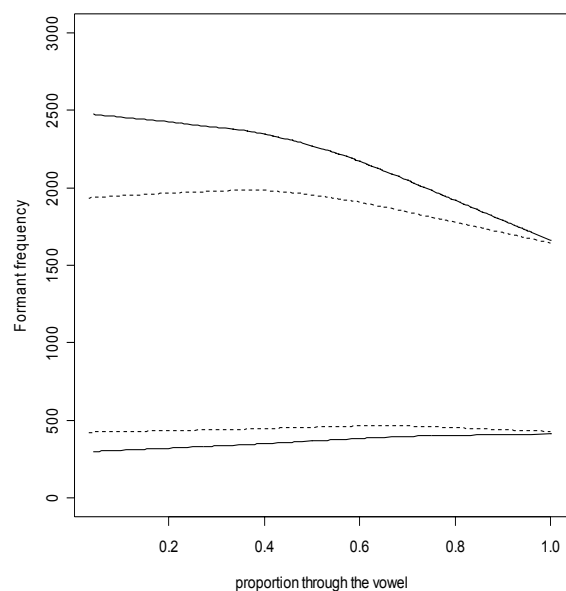


fig. 4.6 (a) – Early males

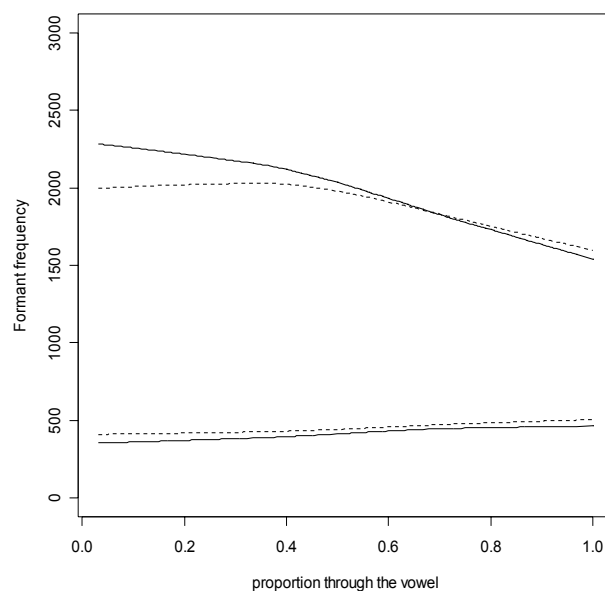


fig. 4.6 (b) – Early Females

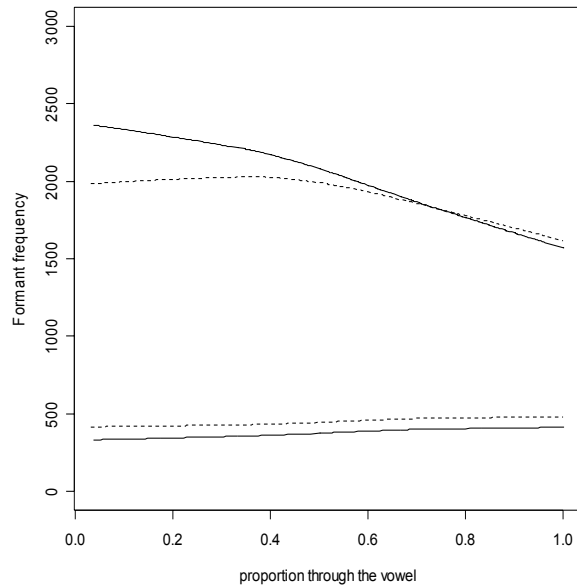


fig. 4.6 (c) – Medium Males

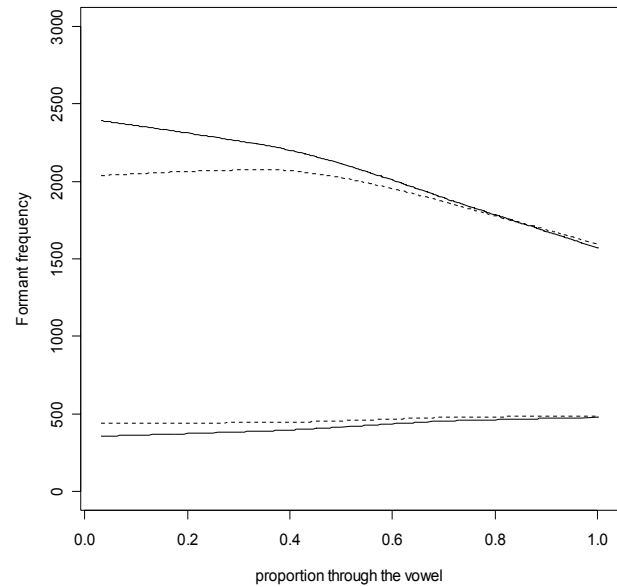


fig. 4.6 (d) – Medium Females

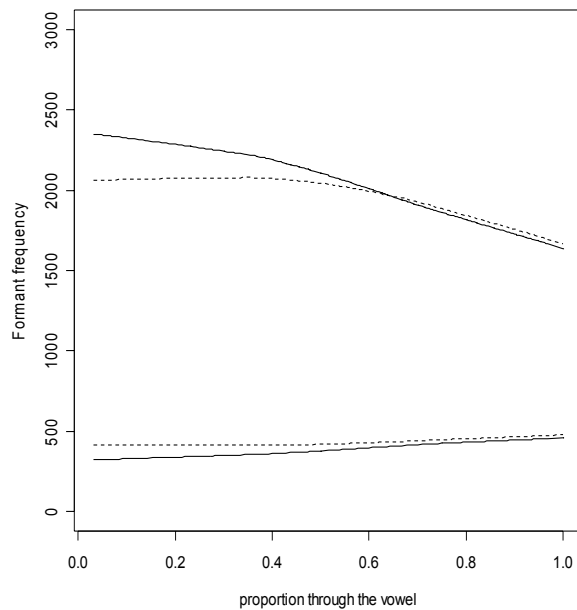


fig. 4.6 (e) – Late Males

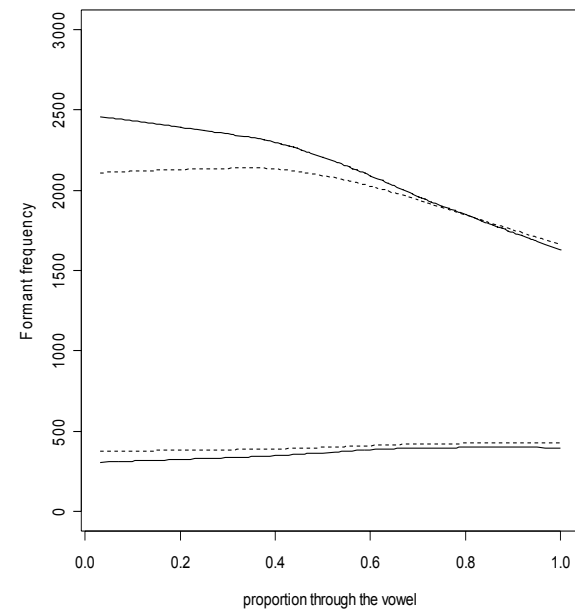


fig. 4.6 (f) – Late Females

Figure 4.6 - Lowess line scatterplot smoother models of NEAR (solid lines) and SQUARE (dashed lines) trajectories in the speech of 6 groups of speakers from the Intermediate Archive. Males are plotted in the left column, early speakers in the top row.

It can be observed that although the trajectories in figure 4.6 support the above conclusions that the merger of the front centring diphthong was well under way during the Intermediate Period, the 'partitioning' of speaker groups seems to be somewhat different. That

is, whereas we have concluded above that the merger gains momentum in the late female group, the trajectories suggest a sub-grouping along the lines of early males vs. all others. More specifically, it is in that group where the trajectories look as one would expect from a dialect that keeps the two vowel phonemes apart. Both formants are distinct along the entire trajectory, whereas there is overlap in one or both dimensions in all the other groups. The medium males keep the first formants of both vowels distinct over the entire trajectory, but show overlap in F2 toward the offglide. The late male group shows the same kind of F2 overlap (if somewhat earlier in the trajectory) as well as approximating F1 contours toward the offglide.

As for the females, all three groups have some degree of overlap in the F2 contour toward the offglide as well as closely bundled F1 contours over the entire trajectory. In keeping with current notions in sociolinguistics that males generally show more conservative vowel realisations in sound change from below than females of the same age, we could therefore justifiably date the precursors of the NEAR-SQUARE merger (if we accept overlap in the offglide as a precursor to the modern pattern of complete overlap) further back than the nucleus analysis suggests.

### *Properties of the offglide*

Having identified the offglide as the locus of early overlap between the front centring diphthongs, we will now turn to a more detailed discussion regarding offglide position and target specification. Since offglides tend to be somewhat weakly specified in comparison to nuclei, it was not always possible to identify comparable positions for point measurements in the offglide. The following analysis is therefore based on the average formant value between 70 % and 90 % through each diphthong token. Hence, 'offglide' here refers to an average value of a number of (between 1 and 6, depending on the overall length of the diphthong) observations. Table 4.3 states the mean formant frequency values as well as the standard deviation and the first and third quartile of the offglide for all six speaker groups. Although there do not seem to exist clear directional patterns over the entirety of the sample, two points are worth noting: In all three groups of male speakers, the standard deviation of the formant frequency values is consistently higher in SQUARE than in NEAR, which may suggest that the

former is less strictly specified for a specific phonetic target than the latter for those speakers. Alternatively, this pattern may arise as a result of greater variation across individuals. Although averages of individual offglide standard deviations indicate that the standard deviation is indeed higher in SQUARE (78.6 hz for NEAR, 96.9 hz for SQUARE) compared to NEAR, individual variance renders this difference insignificant (as determined by a Wilcox test,  $W=46.5$ ,  $p>.2$  in F1 and  $W=60$ ,  $p>.7$  in F2)

Secondly, the highest degree of freedom in the offglide realisation of both vowels seems to hold in the speech of the late speakers, although not in a very straightforward way: Whereas the late females have the highest STDevs in the F2 dimension, the late males allow for a higher degree of freedom in F1.

	<i>E Ma N</i>		<i>E Ma SQ</i>			<i>E F N</i>		<i>E F SQ</i>	
	F1	F2	F1	F2		F1	F2	F1	F2
<b>sd</b>	65.6	201	99.7	250.8	<b>sd</b>	108	223	84.5	231.5
<b>mean</b>	417.9	1921	491	1803	<b>mean</b>	475.7	1736	493.7	1751
<b>1stQ</b>	382.8	1810	430.6	1640	<b>1stQ</b>	392.4	1561	428.9	1597
<b>3ndQ</b>	445.5	2065	526	1920	<b>3ndQ</b>	549.6	1860	558.3	1886
	<i>M Ma N</i>		<i>M Ma SQ</i>			<i>M F N</i>		<i>M F SQ</i>	
	F1	F2	F1	F2		F1	F2	F1	F2
<b>sd</b>	74.6	223.1	90.19	229	<b>sd</b>	89	183.4	96	206.1
<b>mean</b>	430.1	1745	494.2	1790	<b>mean</b>	481.8	1764	509.4	1781
<b>1stQ</b>	381.8	1595	433.5	1617	<b>1stQ</b>	422.7	1664	443	1633
<b>3ndQ</b>	456.8	1880	543.8	1926	<b>3ndQ</b>	543.7	1880	568.7	1924
	<i>L Ma N</i>		<i>L Ma SQ</i>		F	<i>L F N</i>		<i>L F SQ</i>	
	F1	F2	F1	F2		F1	F2	1	F2
<b>sd</b>	118.6	195.6	144.1	216	<b>sd</b>	105.5	259.5	109.7	298.5
<b>mean</b>	469.9	1808	493.9	1859	<b>mean</b>	420.6	1804	459.3	1836
<b>1stQ</b>	397.3	1701	404.8	1698	<b>1stQ</b>	369.7	1702	379.9	1638
<b>3ndQ</b>	519.5	1927	533.7	1859	<b>3ndQ</b>	456.7	1960	514.9	2006

Table 4.3 – Mean formant frequency values, standard deviations and 1<sup>st</sup> and 3<sup>rd</sup> quartiles means in the offglide of NEAR and SQUARE in the speech of six groups of Intermediate speakers.

One might wonder how much importance should be assigned to offglide properties in the first place, given that the most salient property of the merged system in current NZE is overlap in

the nucleus. In addition, the traditional phonemic notation of the offglide as /ə/ in most non-rhotic English dialects indicates that there has not been much of a salient auditory difference in the past between the offglides of /ɪə/ and /ɛə/. Yet, the development in the speech of (especially the male) speakers in this study suggests a rather linear development where the two trajectories are kept apart in the earliest male group, and there is subsequent approximation toward the offglide region of the diphthongs.

#### 4.7 Stress

Finally, we will consider the role of stress in the developments in NEAR and SQUARE over the Intermediate Period. MacLagan and Gordon (1996) stated that stress might have a role to play in the degree of displacement of SQUARE tokens, whereby the higher the degree of stress is, the closer the vowel. Table 4.4 states Hotelling-Lawley scores and significance levels for all speaker groups and stress levels. In section 4.4.1 it was noted that although the late females keep their overall distributions of the two vowels distinct, they do so at a considerably lower level of significance than is the case in the other groups. Additionally, it can be shown that this effect is much stronger if the data is resolved for different levels of stress (stress was assigned on an auditory basis, where 1=highest degree of stress, 2=medium stress and 3=low stress). In comparison to the data given in table 4.1, it can be seen that the front centring diphthong distributions in the speech of the late females are not kept apart at levels of significance below .05 if the data is resolved for stress. In addition, it should be noted that whereas there is at least a trend towards keeping them distinct for tokens of stress level 2 and 3 (i.e. medium and low), sounds in heavily stressed position show a much lower Hotelling-Lawley score and a p-value of .3, which indicates a fairly advanced stage of overlap between these sub-distributions.

A similar pattern holds in the late male group, where the Hotelling-Lawley score is considerably lower in stressed contexts, although the vowels are kept distinct in this group for all stress levels, but at an appreciably lower level of significance in tokens of stress level 1. With respect to the early and medium groups of both sexes, it can be observed that solid distinctions between NEAR and SQUARE are maintained for any level of stress.

Two general points are worth noting: First, for any two comparable sex cohorts, it is the female speakers who tend to show lower levels of differentiation of the two vowel phonemes. This effect is particularly striking in the late sample, but also present in the sub-distributions of vowels of stress level 3 in the early and medium groups. Furthermore, there is a striking mismatch between the absolute distance measures (i.e. the H-L traces) and the significance levels in the early female group, which suggests a rather tight clustering of the two distributions in these groups. This effect, however, as well as the global one mentioned above with respect to significance levels in the late samples is probably in part ascribable to smaller token numbers in those sub-samples. This caveat is also supported by the author's auditory impression, which suggested that although the realisation of SQUARE is rather high in articulatory space in especially the late female group, most tokens were heard 'correctly' (~ 1/6 of all SQUARE tokens were auditorily coded as NEAR in that group, compared to ~ 1/15 for the entire sample including the late female group, and ~ 1/17 excluding that group).

<i>Speaker group</i>	<i>Stress</i>	<i>Hotelling-Lawley trace</i>	<i>p</i>	<i>n</i>
<b>Early females</b>	1	0.1721	< .0001	100
	2	0.3911	< .0001	81
	3	0.3145	< .01	51
<b>Medium females</b>	1	0.6059	< .0001	74
	2	0.823	< .0001	79
	3	0.3277	< .01	50
<b>Late females</b>	1	<b>0.09426</b>	<b>.2964</b>	46
	2	<b>0.2410</b>	<b>.08349</b>	36
	3	<b>0.23539</b>	<b>.08796</b>	27
<b>Early males</b>	1	1.427	< .0001	68
	2	1.306	< .0001	59
	3	2.029	< .0001	55
<b>Medium males</b>	1	0.5442	< .0001	81
	2	0.655	< .0001	100
	3	0.8193	< .0001	51
<b>Late males</b>	1	0.1326	< .05	57
	2	1.0895	< .0001	48
	3	0.9248	< .001	31

Table 4.4 – Hotelling – Lawley scores and significance levels for /ɪə/ - /ɛə/ distributions in the speech of six groups of Intermediate speakers, resolved for three levels of stress. Numbers in bold indicate distributions that are not significantly distinct at  $p < .05$ . Stress levels are stated in descending order prominence, i.e. 1 = heavy stress, 2 = medium stress and 3 = low stress.

#### 4.8 Merger by approximation, transfer, or expansion?

Recall from section 2 above that there are essentially three mechanisms by which a phonemic merger can come about: (1) merger by approximation, where one vowel gradually approximates the vowel space occupied by another vowel; (2) merger by transfer, whereby words containing a given vowel change that vowel to another phoneme until all words that contained the original vowel contain the new vowel; and (3) merger by expansion, where tokens from two originally distinct vowel distributions spread out into each other's vowel space and thereby blur the contrast between the two.

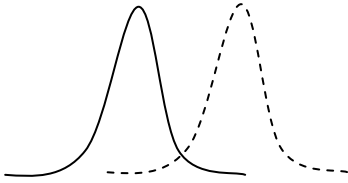
A reasonable way of assessing this in a particular case would be an investigation of the distributional properties of the two vowels in a pre-merger stage. If the merger comes about by approximation, we would expect each of the two vowels to maintain approximately normal distributions throughout the entire process. In terms of pooled data from both vowels, we would expect a bimodal initial stage developing into a normal distribution after the merger is completed (probably via a trimodal in-between stage since the outliers between the nodes will add up in a situation of close proximity of the two vowels). The exact shape of this type of transitional distribution would, however, depend on a number of factors such as the flatness of the individual distributions as well as the number of tokens within each of the individual distributions. If the number of tokens in one of them were greater than in the other, we would expect that mode to be 'swamped' by the transitory middle mode earlier than the mode with the higher token numbers.

As for merger by transfer, we would expect no approximation of distributions, but rather subtraction from one distribution over time since in this scenario, a given word can either contain vowel A or vowel B, but no intermediates. Finally, a merger by expansion would imply an increase in flatness of each of the distributions over time. The scenarios are schematised in figure 4.7.

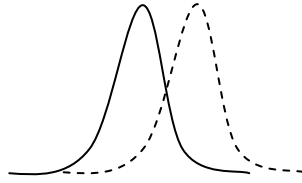


*merger by approximation:*

stage 1 – initial



stage 2 – transitional

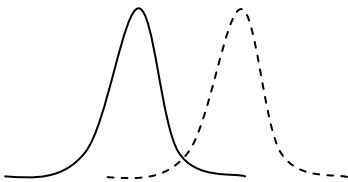


stage 3 -final

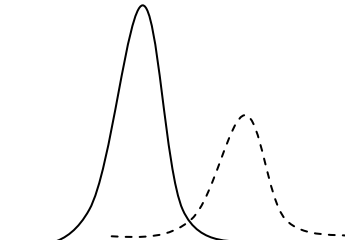


*merger by transfer:*

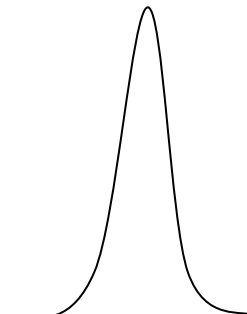
stage 1 – initial



stage 2 – transitional

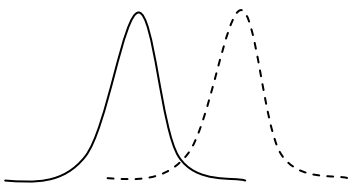


stage 3 - final

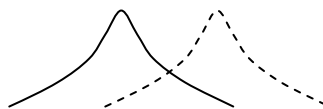


*merger by expansion:*

stage 1 – initial



stage 2 – transitional



stage 3 - final

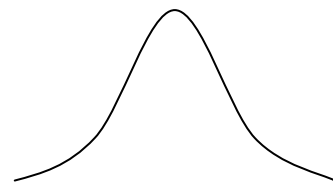


Figure 4.7 – Hypothetical token distributions of two vowels undergoing merger.

Figure 4.8 below plots pooled NEAR/SQUARE data from the Intermediate sample. In addition, figure 4.8 shows the distributions for each of the vowels, respectively.

### Intermediate speakers - NEAR + SQUARE F1 distribution

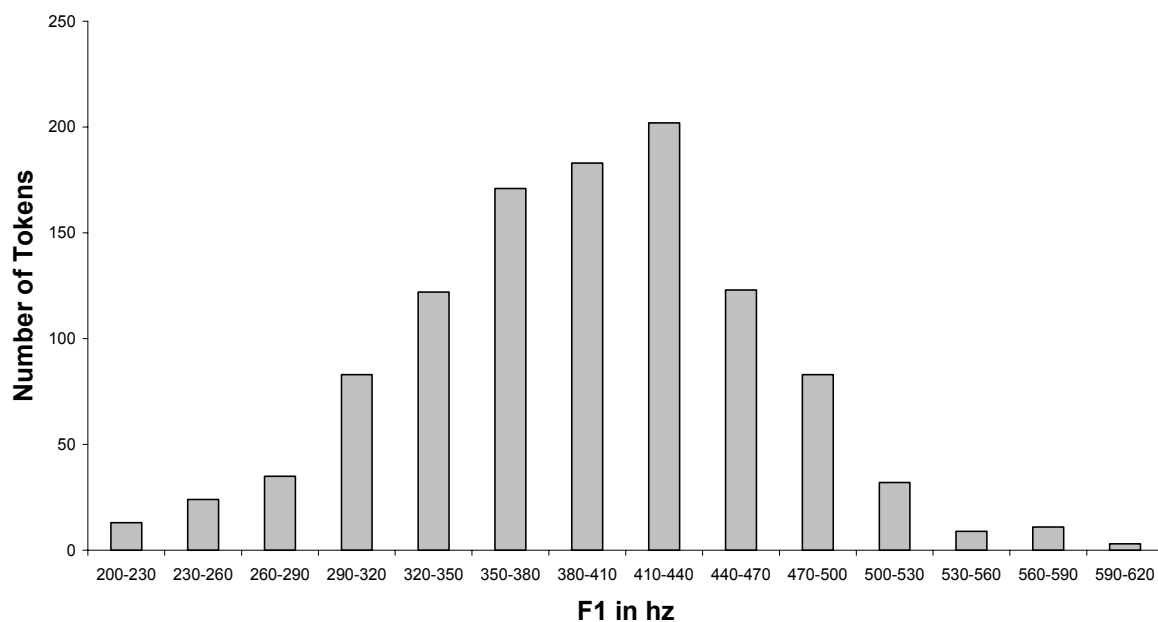


Figure 4.8 (a)

### Intermediate Speakers - NEAR F1 distribution

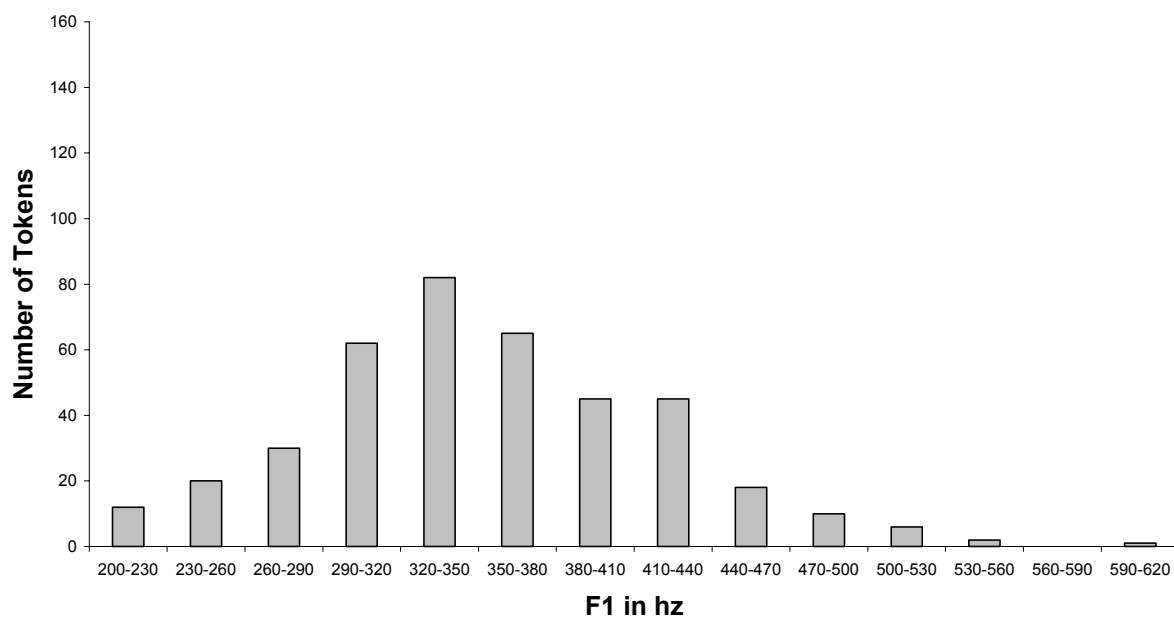


Figure 4.8 (b)

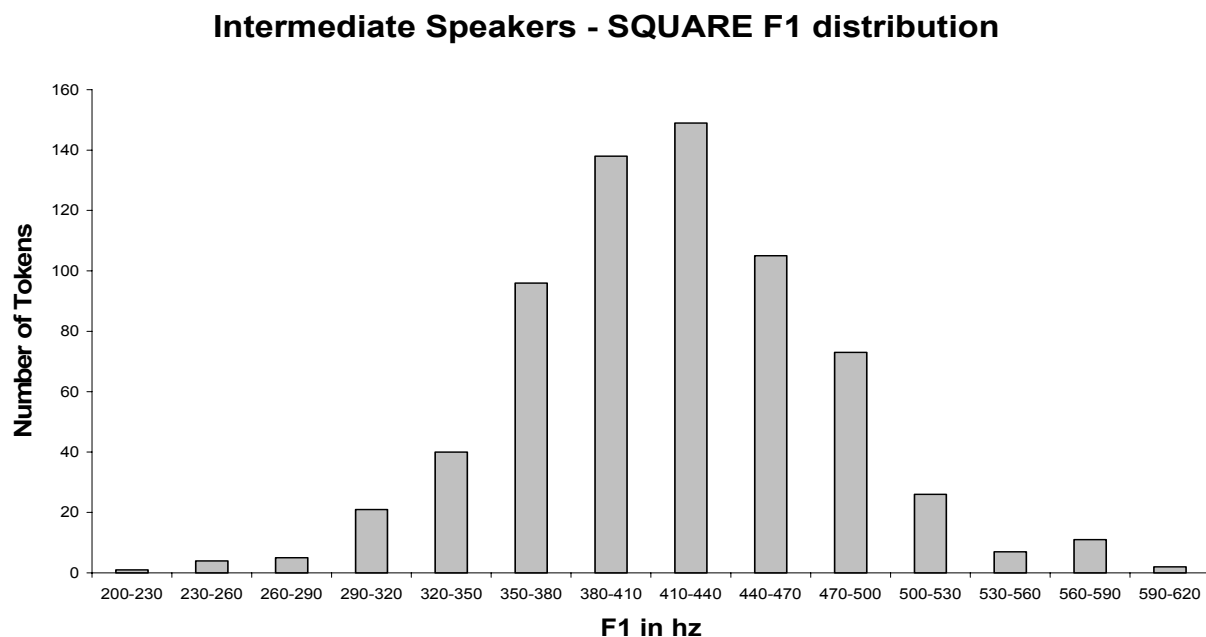


Figure 4.8 (c)

Figure 4.8 – Token distributions of pooled F1 data from the Intermediate sample. 4.8 (a) shows the entire data pool, 4.8 (b) shows NEAR, 4.8 (c) shows SQUARE.

Figure 4.8 (a) shows a slightly left-skewed unimodal distribution, which indicates that in the Intermediate period as a whole, the two vowels are closer to each other than would be expected between vowels that do not overlap in the F1 dimension. Recall from the discussion above that a unimodal distribution can come about in two ways: Either as a ‘truly’ normal distribution, i.e. what is plotted are tokens of the same category, or as a ‘hidden bimodal’ distribution, whereby two unimodal distributions are sufficiently close to each other that the sum of their respective outlying tokens exceeds that of the mean of either distribution. Figures 4.8 (b) and (c) suggest that the latter is the case: Both NEAR and SQUARE show unimodal distributions. However, two properties stand out in comparison of the two vowels. First, token numbers in NEAR are somewhat smaller than in SQUARE. In addition, the NEAR distribution seems somewhat flatter than SQUARE. This, in addition to the relative proximity of the means, brings about the overall unimodal left-skewed distribution. We are therefore faced with a somewhat odd situation: Whereas the F1 distributions of NEAR and SQUARE, if looked at in separation, are clearly distinct from each other (cf. the H-L trace score calculations above), the overall pooled F1 data results in a non-distinct distribution.

In addition, the lack of bimodality in either of the two vowels strongly suggests that we can exclude a merger-by-transfer type of scenario. What is less clear is whether we can restrict the NEAR/ SQUARE merger to either one of the remaining types, i.e. merger by approximation or by expansion. The data given in table 4.1 above suggests that we might be faced with a combination of two processes. If we compare the early male and the late female group (cf. table 4.2), it appears as if there is evidence for both processes. That is to say, whereas the position of the nucleus of SQUARE is closer to that of NEAR in the late female sample in the F1 dimension, the SD of NEAR is clearly higher in the late female sample as well. Figure 4.9 below plots all NEAR/SQUARE tokens from each respective speaker sample. The lines outline an area 2 standard deviations from the mean in each dimension.

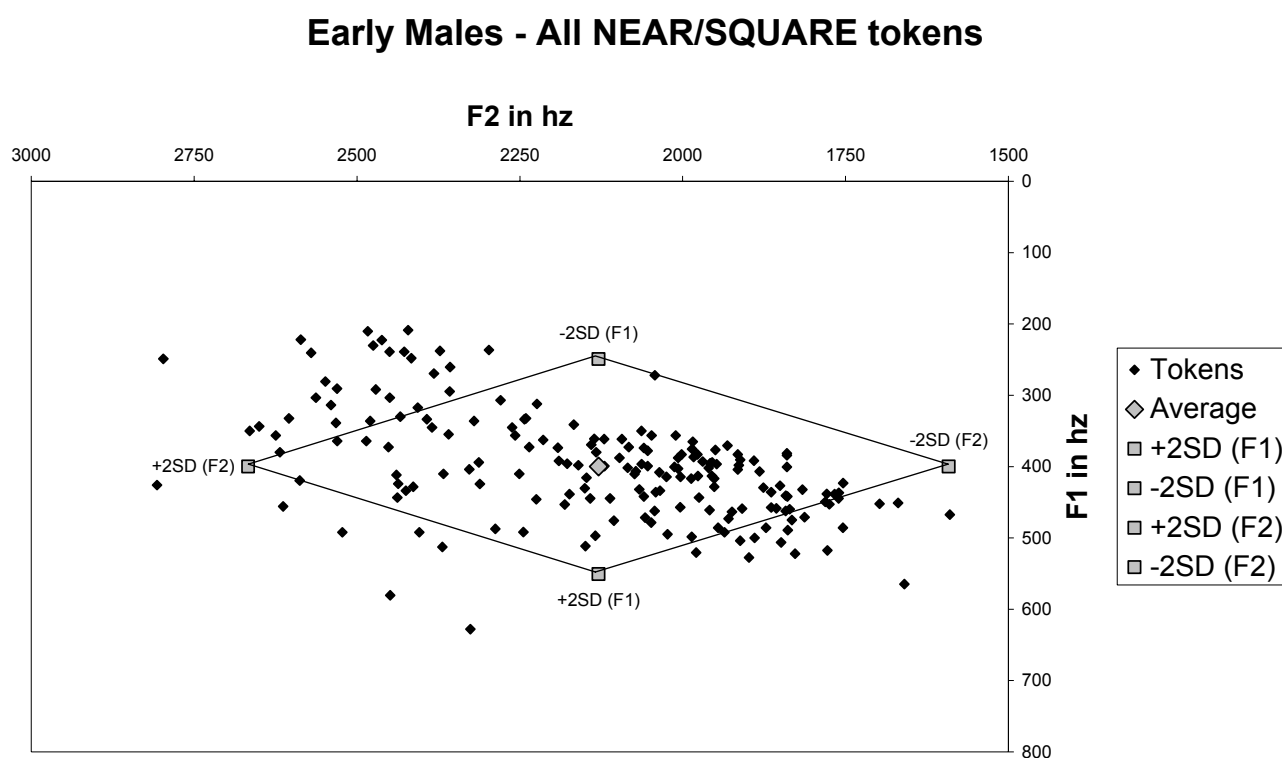


Figure 4.9 (a)

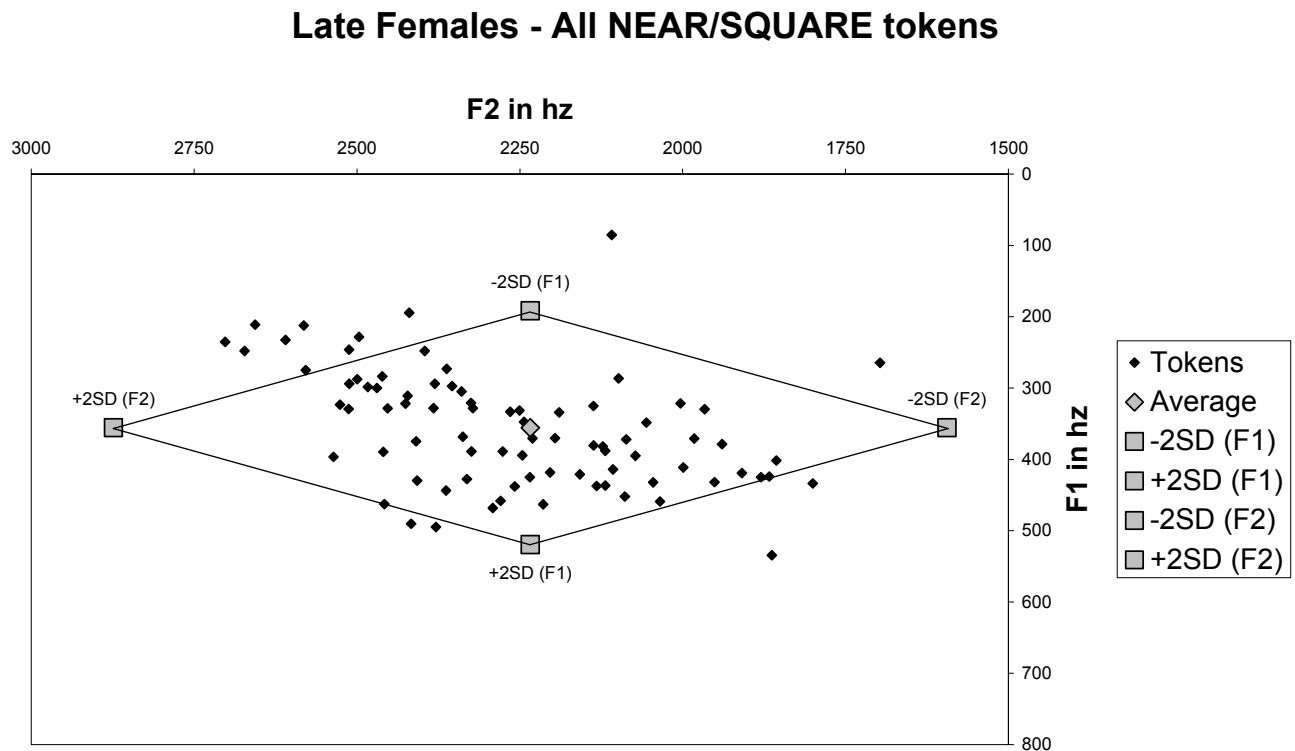


Figure 4.9 (b)

Figure 4.9 – Token distributions of NEAR/SQUARE in the early male (4.9 (a)) and the late female (4.9 (b)) sample. The outlined area around the mean delineates 2 standard deviations from the mean.

We observe in figure 4.9 that in comparison to the early male sample, the late females show both an upward/leftward shift of the overall distribution of tokens as well as a larger standard deviation from the mean in pooled data. Again this suggests that the space available for the two phonemes does not indeed become smaller as would be expected if the merger were purely one of approximation. To some extent, however, the leftward displacement (i.e. toward the direction of NEAR) in the late female sample seems to be carried by smaller token numbers of SQUARE compared to the early male sample: Whereas in the overall data pool of the early male NEAR/SQUARE sample we find a ratio of 58 (NEAR) to 124 (SQUARE), or 1: 2.13, the ratio is 32/50 in the late female sample, which is 1:1.56. On the other hand, the data shown in table 4.5 indicates that this effect is not solely an artefact of skew in token numbers, since the late female SQUARE vowel is closer to NEAR in both formant frequency dimensions compared to the early males, whereas in NEAR, the values are approximately similar in both samples. It should also be noted that the standard deviations themselves do not seem to be

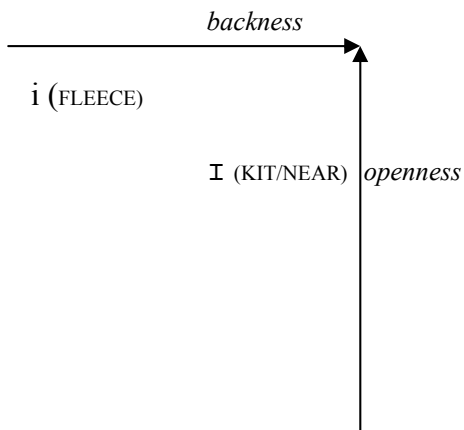
mainly due to the smaller token numbers in the late female samples. Ten random samplings of 82 tokens (i.e. the number of tokens observed in the late female sample) from ten random permutations of the early male sample yielded an average standard deviation of 149.1 hz (F1)/563.87 hz (F2), which is more similar in magnitude to the standard deviations found in the early male sample than to those in the late female one.

Overall, a picture of the NEAR/SQUARE merger emerges which seems rather less tidy than idealised histories, or schemata, would lead us to believe. In general, the conclusion seems valid that the merger is in the direction of NEAR, and that it is a merger-of-approximation type of process. However, we note at least the following further points:

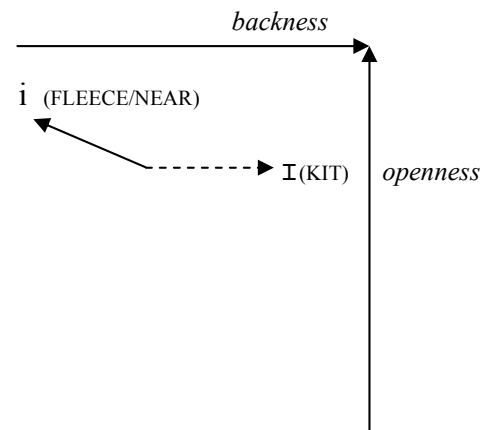
- (1) The nucleus of NEAR is anything but stable. First of all, there seems to be a transitional process whereby NEAR reacts to raised SQUARE by raising/fronting itself. In addition, there is some evidence that the vowel spaces occupied by both NEAR and SQUARE are larger in the later samples, although caution is advised with regard to this statement due to different token numbers across the subsamples as well as the non-categorical nature of this relation (i.e. the expansion of the vowel space in NEAR is evidenced only in the F1 dimension, where both late males and females show a larger SD; cf. table 4.2 As for SQUARE, the picture is more confusing since the late females show the largest SD around the mean in F2, whereas the late males show the smallest). If this is actually the case, we would have to conclude that the NEAR / SQUARE merger in NZE shows properties of both a merger-by-approximation as well as one of expansion. Whereas this may well be true, there is one further complication which allows for an alternative explanation: As was pointed out in the introductory section above, there seems to be agreement that at some stage after the GVS, the nucleus of NEAR became associated with the KIT vowel, and that this was the original quality in NZE (as well as its closest varieties). The fact that the further developments in KIT do not affect the nucleus of NEAR in NZE (i.e. the latter does not undergo centralisation) has been suggested to be indicative of an association of the NEAR nucleus with FLEECE. As far as I am aware no one has made any suggestion as to whether this was a gradual type of process or a discrete transfer across categories. If the former was the case, the NEAR nucleus would have undergone a process of

raising/fronting in NZE *independently of any process affecting SQUARE*. This is sketched below:

stage 1  
(initial, where NEAR = KIT)



stage 2  
(final, where NEAR = FLEECE)



Given the assumption that this was a gradual process, and that the nucleus of NEAR is now FLEECE whereas it was KIT at the earliest stages of NZE (this is also the impression of Margaret MacLagan upon inspection of the ONZE data; personal communication), this would leave us with a rather restricted timeframe for the process, namely between late MU/early intermediate and contemporary NZE. We may therefore have to allow for a scenario whereby the raising/fronting of NEAR and SQUARE are independent processes.

A more in-depth analysis of this will be the topic of chapter 5.

- (2) The distinction between merged and non-merged vowels is blurred, at least if we take into consideration the properties of token distributions. Although in all six subsamples analysed here, the distributions of NEAR and SQUARE are distinct, they are also sufficiently close to each other to show a non-distinct (i.e. unimodal) overall distribution. Therefore, the set-up of the two vowels in the Intermediate period is probably best regarded as a pre-merged stage with distributional properties that are different from both non-merged and fully merged vowel distributions. Another way of putting this would be to call it a 'near-merger'. It is unclear whether Intermediate speakers actually perceived the distinction between NEAR and SQUARE, and the fact that they maintain it does not necessarily imply

that they are aware of it. In light of recent exemplar-dynamic approaches to phonological development, it appears questionable whether a language learner who is faced with token distributions of the type shown in figure 4.9 above can actually infer two distinct phonemic types. The statement that I have made rather confidently throughout this section, as well as earlier (Langstrof 2004), as to the clear distinctness of NEAR and SQUARE in the Intermediate period if assessed auditorily may well be an artefact of my own non-NZE categorisation as well as phonetic training<sup>31</sup>.

#### 4.9 Lexical frequency

In this section we will briefly look at the role of lexical frequency in the distribution of formant frequency values in the FCD data. Table 4.6 below shows correlation coefficients and significance values for F1/F2 – lexical frequency in NEAR/ SQUARE.

		<i>word frequency</i>		<i>lemma frequency</i>	
		<i>rho</i>	<i>p</i>	<i>rho</i>	<i>p</i>
NEAR	F1	<b>-.16</b>	<b>&lt;.01</b>	<b>-.25</b>	<b>&lt;.0001</b>
	F2	<b>.27</b>	<b>&lt;.0001</b>	<b>.28</b>	<b>&lt;.0001</b>
SQUARE	F1	.05	> .2	.05	> .2
	F2	.04	> .2	.06	> .15

Table 4.6 – Correlation coefficients and significance levels between formant frequency and lexical frequency.

It appears that while there is no correlation between formant frequency and lexical frequency in SQUARE, solid correlations hold for both formants in NEAR. More specifically, it seems clear that more frequent words show fronter/closer realisations than infrequent ones. However, there are reasons to believe that this effect is entirely artefactual. Figure 4.10 below plots word frequency against each formant, respectively.

<sup>31</sup> There is some anecdotal evidence for this: I occasionally cross-checked my auditory coding with a native speaker of American English in cases where I was uncertain whether a given token was NEAR or SQUARE. This mainly concerned etymological SQUARE tokens that seemed to me ambiguous between the two. To said American colleague, these sounded like NEAR without exception.



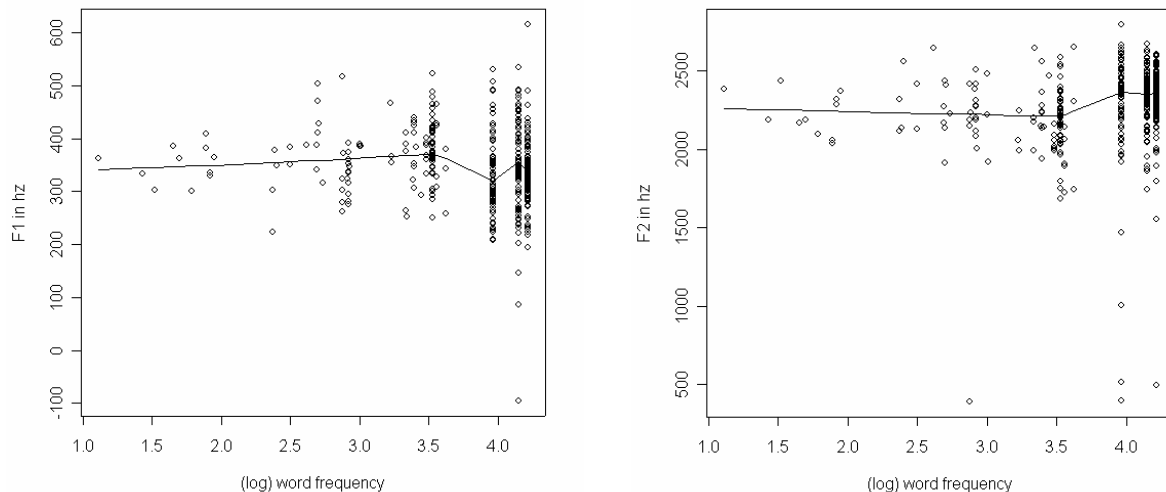


Figure 4.10 – Formant frequency plotted against lexical frequency in the Intermediate sample. The lines represent locally weighted scatterplot smoothers through the data.

It appears as if the entire effect was carried by two points on the x-axis in 4.10 (a), namely the rightmost one and the third from the right, as well as the three rightmost ones in the F2 dimension (4.10 (b)). Apart from these words being comparably common, they also occur frequently in the sample under analysis here, which solidifies the overall correlation. The lexical items in question are *year* (log frequency 3.96), *here* (log frequency 4.15) and *years* (log frequency 4.21). That is, the frequency effect in the f1 dimension is probably due to phonetic factors, i.e. the tokens that carry the effect both occur in both frequent words as well as in a consonantal environment which strongly favours the observed effect, i.e. post-palatal. If the two lexical items are removed from the data pool, the correlation between lexical frequency and F1 no longer holds ( $\rho = -.002$  ;  $p > .8$ ). Similarly, the correlation between F2 and lexical frequency rests on the strength of only a small number of lexical items. However, if we remove the post-palatal *year* and *years*, the correlation remains significant at  $p < .001$ . Figure 4.11 below plots word frequency against F2 for NEAR, excluding all tokens of *year* and *years*.

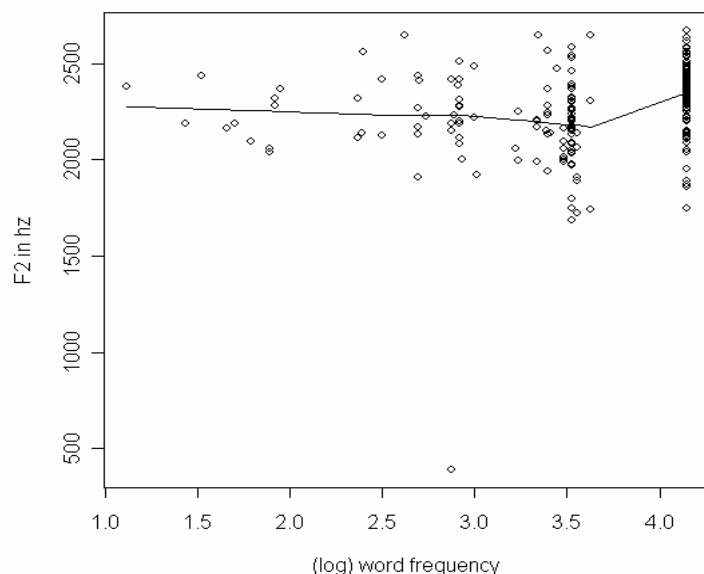


Figure 4.11 – Formant frequency plotted against lexical frequency for NEAR in the Intermediate sample. The lines represent locally weighted scatterplot smoothers through the data. All occurrences of the words *year* and *years* are removed.

It is obvious that the effect is carried by the lexical item *here*. Although phonetic factors cannot account for this effect (i.e. the environment /h\_ə#/ is as neutral an environment as one can obtain), *here* is the only lexical item within the NEAR data that comes close to what one might call a function word, and it has been pointed out before that frequency effects may, in some cases at least, well be artefacts of lexical category effects whereby function words behave differently from content words, a claim that is well supported by the present data (although the reasons are less clear in this particular case: Why would a frequent word, i.e. a word whose vowel identity can be more easily inferred by its overall shape, spearhead a transitional change that may be regarded as a process of chain-shift-like properties, i.e. increasing distinctness to an encroaching vowel SQUARE?)<sup>32</sup>.

<sup>32</sup> Incidentally, this particular case also casts doubts onto accounts of chain shifts and mergers that call upon ‘functional load’ effects. In brief: Why would a lexical item that has little potential of being confused with a SQUARE word be one of the furthest from the distribution of SQUARE?

#### 4.10 Durational Properties of the FCDs

Having looked at the formant frequency properties of NEAR and SQUARE, we will now turn to considering the duration of the front centring diphthongs. As was argued in chapter 3 above, there are solid durational differences between the short front vowels than cannot be ascribed solely to mechanical factors, and which play a role in the analysis of distance relationships between phonemes in chain shift processes (cf. also chapter 6 below). Analysing the durational properties of the front centring diphthongs might therefore be important to obtain a better picture regarding the relationship between the monophthong and the diphthong system. More specifically, one might wonder whether there is any justification (beyond etymology) to equate the nuclei of certain types of diphthongs (NEAR and SQUARE in this case) with elements from the monophthong system, and what these equivalences are. More specifically, if the nuclei of the front centring diphthongs are in some meaningful sense ‘the same’ as the short front vowels, one would probably expect similar relationships between the elements of the front diphthong set and those of the monophthong set along all phonetic dimensions, including duration (cf. chapter 5 for an analysis of this assumption).

Table 4.7 sums up the duration measurements for the two centring diphthongs for the two speaker groups considered above (early males and late females). Note that the data presented represents the average of all NEAR/SQUARE tokens for each speaker, irrespective of phonological environment.

<i>mono</i>				<i>poly</i>			
Vowel	speakers	env.	duration	Vowel	speakers	env.	duration
<b>NEAR</b>	<b>BOTH</b>	all	175.7	<b>NEAR</b>	<b>BOTH</b>	all	120
		_[+voice]	166.5			_[+voice]	120.1
		_-voice]	150.9			_-voice]	n/a (n=0)
		_#	186.2			_#	n/a (n=1)
	<b>Early Males</b>	all	175		<b>Early Males</b>	all	126
		_[+voice]	165.8			_[+voice]	126
		_-voice]	n=0			_-voice]	n/a (n=0)
		_#	185.2			_#	n/a (n=0)
	<b>Late Females</b>	all	171		<b>Late Females</b>	all	103.4
		_[+voice]	155.4			_[+voice]	103.4
		_-voice]	173.2 (n=3)			_-voice]	n/a (n=0)
		_#	184.2			_#	n/a (n=0)
<b>SQUARE</b>	<b>BOTH</b>	all	164.8	<b>SQUARE</b>	<b>BOTH</b>	all	134.3
		_[+voice]	151.1			_[+voice]	136.2
		_-voice]	137.6			_-voice]	107.7
		_#	171.9			_#	144
	<b>Early Males</b>	all	166.7		<b>Early Males</b>	all	134.9
		_[+voice]	169.2			_[+voice]	140.1
		_-voice]	145			_-voice]	105.5
		_#	166.5			_#	n/a (n=0)
	<b>Late Females</b>	all	153.6		<b>Late Females</b>	all	118.1
		_[+voice]	119.1			_[+voice]	118.1
		_-voice]	159.2			_-voice]	n/a (n=1)
		_#	170.4			_#	n/a (n=0)

Table 4.7 - Duration averages of the front centring diphthongs in two Intermediate speaker groups (early male and late female speakers). The data lists monosyllables and polysyllables separately. The phonological environments listed are \_[+/-voice] (i.e. followed by a voiced/voiceless consonant, respectively) and \_# (word-final position). Token numbers are indicated for  $n < 5$ .

#### *Centring diphthong duration in monosyllables and polysyllables*

Overall, the NEAR diphthong is slightly longer than SQUARE. If we consider all monosyllables, the difference is significant only in the late group ( $W = 3102$ ;  $p < .05$ ). In addition, this effect seems to be carried mainly by tokens before voiced consonants ( $W = 340$ ;  $p < .01$ ). Again, this seems to violate the expectation that if we compare two vowels of the same ‘intrinsic length’ (cf. the discussion of vowel duration in chapter 3 above), the closer one would be shorter. In polysyllables, there is a significant overall duration difference between the two centring

diphthongs ( $W = 3693$ ,  $p < .05$ ). However, the difference is the inverse of what was found above for monosyllables in that here, SQUARE is longer than NEAR. If we break down the data into subsamples, the statistical significance vanishes (presumably due to low token numbers within most of the cells). Across age groups within the same vowel, only the difference in SQUARE is significant ( $W = 972$ ;  $p < .05$ ), i.e. late speakers have a shorter SQUARE vowels in polysyllables than do early speakers.

### CART

Figure 4.12 shows a CART model on pooled data from the early and late speakers (where duration is the dependent variable regressed on F1, F2, age (*early/late*), syllabicity (*mono/poly*), and voicing of the following consonant (*voiced/voiceless/final*)).

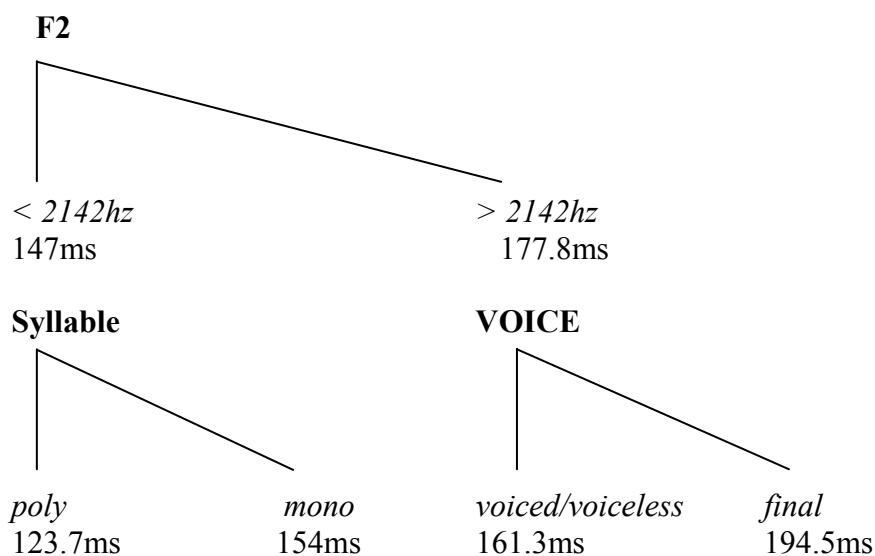


Figure 4.12 – Classification and regression tree analysis of duration in centring diphthongs along the variables *F1/F2/age/syllable/voice*.

Figure 4.12 shows that the best predictor of duration in the centring diphthongs is the second formant frequency. The correlation is significant ( $\rho = .265$ ,  $p < .0001$ ) and holds in both NEAR and SQUARE. Within the group of tokens of shorter duration (i.e. lower F2), a further division is found where vowels in monosyllables are longer than those in polysyllables by a ratio of 1.24/1. Within the group that has higher F2 values/duration, the subordinate division is in terms of the variable voice. Interestingly, the expected duration difference of vowels

depending on whether they occur before a voiced or a voiceless segment turns out to be less significant than whether the vowel is followed by any consonant or not. However, it should be pointed out that in the NEAR/SQUARE data, a ‘following voiced consonant’ is usually the liquid [ɹ], an environment which seems to be exempt from the ‘lengthening 2’ process described in Lass (1992, cf. also chapter 6 of this thesis), where vowels underwent lengthening before voiced stops and nasals.

What seems to more noteworthy in the case of the centring diphthongs vis-à-vis the SFV data analysed above is the complete absence of a length distinction between the two vowels, which suggests that whatever their intrinsic length, it is probably the same in either case. The implications of this are as follows: Although it seems that there is a significant difference in duration between NEAR and SQUARE in the speech of the late speakers, this difference is not nearly sufficient in order to resolve overlap between the two distributions: Figure 4.13 plots duration against F1 in the late sample, and the Hotelling-Lawley trace score calculated for the two distributions along the factors F1/duration is .09 ( $F=8.4$ , 79 df,  $p = .03$ ), which is slightly lower than the one given in table 4.1 above for the same tokens in F1/F2 space. This implies that duration is not of much help in distinguishing NEAR and SQUARE distributions. On the whole then, it seems as if variation in vowel duration is a property of speakers rather than the vowels themselves in the case of the centring diphthongs<sup>33</sup>.

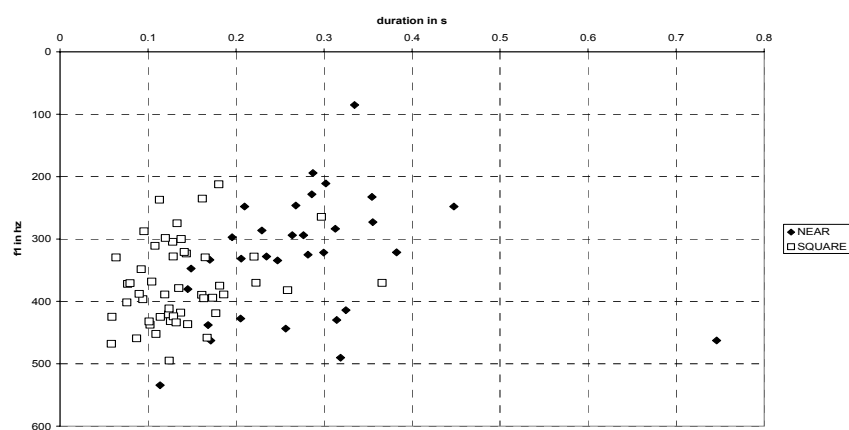


Figure 4.13 – NEAR/SQUARE distributions of 4 late female speakers, with duration (in seconds) on the x-axis against the first formant frequency (F1) on the y-axis.

<sup>33</sup> It has to be borne in mind that in the above discussion, normalised formant frequency values are compared to raw duration measurements, which might obscure those factors that are known to affect duration, but are not structurally motivated (such as speech rate).

#### 4.11 Conclusion

This chapter presented an analysis of the front centring diphthongs NEAR and SQUARE in the intermediate period of New Zealand English. In section 4.3 it was shown that SQUARE undergoes raising over that period of time, and that there also appears to be an upward movement of NEAR, albeit to a lesser extent than is the case in SQUARE. Section 4.4 presented a CART factor analysis of F-distribution, where it was shown that along most dimensions, vowel identity is not a good predictor of the distribution of F-values. Rather, these distributions are best predicted by contextual effects, which in turn suggests that we are faced with a pre-merger state between the two vowels within the Intermediate sample. A trajectory analysis (section 4.6) revealed that it is only in the most conservative sample (i.e. the early males) that we find separate formant frequency contours between the two vowels along the entire trajectory. All other groups show at least some degree of cross-over between the F-contours of NEAR and SQUARE. It was also shown that the offglide of the SQUARE diphthong is less strictly specified than that of NEAR, and that this trend seems to increase over time. An analysis of the approximation of the two vowels under different levels of stress (section 4.7) revealed that (a) late speakers show significantly less distinct distributions of tokens that have *high* stress, and that (b) the difference between the distributions fails to reach significance in the late female sample, which furthermore corroborates the finding that the loss of contrast between NEAR and SQUARE dates back to at least the intermediate period.

In section 4.8, the distributional properties of the front centring diphthongs were analysed further, and it was shown that a scenario where lexical items are transferred from one vowel to the other (i.e. merger by transfer) can be ruled out. Rather, the overall process involves mainly approximation of the two phonemes, with the additional possibility of expansion of the overall vowel space occupied by the two phonemes. It was argued in section 9 that lexical frequency effects do not play much of a role in the merger of NEAR and SQUARE. In addition, the strongest effect (i.e. the F2 effect in *year/years*) may justifiably be ascribed to phonetic factors.

It can therefore be concluded that the developments affecting the front centring diphthongs in NZE involve a complex combination of sub-processes, namely (a) the raising of the nucleus of SQUARE, (b) incipient reactionary raising of the nucleus of NEAR, (c)

eventual merger of the two. In addition, it was argued that even the earliest speakers in the intermediate period show what can be described as a ‘pre-merger’ stage whereby the overall distribution of the two diphthongs, albeit still statistically different from each other in F1/F2 space, shows one prominent distributional mode, which might well be the breaking point where language learners infer a single distribution and thereby acquire a truly merged system. Evidence that the merged state seems to be present in the speech of the late speakers comes from an analysis of the realisation of NEAR and SQUARE under different degrees of stress (cf. section 4.7), where late speakers show a smaller distance between the two nuclei in contexts of high stress. If the two vowels were truly different, one would presumably expect the opposite pattern.

This also raises a problem in how to date back a merger, i.e. has it ‘occurred’ only if there is no difference in either production or perception between the merged phonemes? Or of that difference holds only in certain linguistic contexts (such as low stress)? Does one look at the shape of the overall distribution or at the degree of significance of the differences between the two distributions? It is a statistical verity rather than an empirical finding that, once one vowel distribution approaches another one, there holds an intermediate stage where a unimodal distribution arises due to the adding up of outliers at opposite ends of each of the distributions in question<sup>34</sup>, while the two distributions are still statistically different. Therefore, different ways of looking at the data may be mutually exclusive, and it is not clear to me which one is the more meaningful one. Describing a linguistic situation as holds in the front centring diphthong system in the intermediate period as a ‘pre-merger’ stage is therefore not particularly insightful, and more research into the perception and production of vowel distributions (as well as meta-discussions on this point) is required in order to satisfactorily answer these points.

---

<sup>34</sup> This is a statistical verity only if one realises the fact of variance in vowel token distributions and chooses an analytical approach that incorporates that variance, such as formant frequency measurements and their plotting as populations in F1/F2 space.



## 5.

What is a FCD nucleus? – *On the (non)-correspondence between the nuclei of the front centring diphthongs and their monophthongal (non)-counterparts.*

## 5.1 Introduction

This chapter further explores the relationship between the front centring diphthongs and the KIT/DRESS vowels. More specifically, we will address the following question: If a mid-high monophthong (DRESS) pushes out an adjacent segment of the same type (KIT) in a push chain, why does it fail to do so in another subsystem (the centring diphthongs, where the two eventually merge)?

Stated as such, this question makes a number of assumptions, which can be summed up as follows:

- (1) The nuclei of NEAR/SQUARE are of the same type as are those of KIT/DRESS. That is, whether we regard the nuclei of the centring diphthongs as either long ('tense') or short ('lax'), they are the same as the SFVs.
- (2) Within each of the two subsets, we are dealing with 'the same' elements. That is, DRESS and KIT are in the same relationship to each other as are SQUARE and NEAR.
- (3) In combination, this implies that (a) NEAR is to KIT as SQUARE is to DRESS and (b) NEAR is to SQUARE as KIT is to DRESS.
- (4) More generally, the question implies that diphthongal nuclei have a parallel element in the monophthong system, i.e. they are not 'something different'. Since this point may not be obvious, it requires a brief discussion. All dictionaries of English that I am aware of transcribe the nuclei of centring diphthongs (in fact, any diphthong) with

some symbol that otherwise represents a monophthong. The question is whether this is simply transcriptional convenience, or whether there are good reasons for assuming this type of equivalence. In the case at hand, the primary justification seems to be historical: The FCDs originate as pre-rhotic allophones of monophthongs (ME /e:/, /ɛ:/, /a:/ and /ai/). That is, they were instantiations of the ‘same thing’ (at least phonemically) at some stage in the history of English. In addition, Wells (1982:153) gives an example of a phonological alternation involving an FCD and a monophthongal vowel. The example is: *severe* /sɪvɪ(i?)ə/ <-> *severity* /səvɛrɪtɪ/, i.e. an instantiation of an alternation that stems from trisyllabic laxing, whereby long vowels correspond to short vowels in certain derivational paradigms (cf. the famed *divine-divinity*, *serene-serenity*, Chomsky and Halle (1968)). The ‘rule’ is roughly as follows: In deriving a word by adding two unstressed syllables from a word that contains a stressed long vowel (this normally involves deadjectival nouns in *-ity*), the stressed root vowel is shortened and lowered by one slot in the derived word. If the root vowel is an upgliding diphthong, the derived word has a short high vowel. This of course mirrors the historical development whereby only long vowels undergo upward shift by one slot (the GVS). Now if the alternations in *severe-severity* and *serene-serenity* are in some sense equivalent, there would be some justification in extrapolating the same element into the underlying form of *severe* as is the case in *serene*. Assuming that monophthongs are in some sense less ‘marked’ than diphthongs (or in any case more frequent in this particular case), it would seem prudent to take the monophthong representation as the basic one. We therefore seem to have some phonological as well as historical reason for assuming correspondence between diphthongal nuclei and elements from the monophthong system. Note that although this has (primarily) a bearing on the question of whether there are *any* structural correspondences between diphthongs and monophthongs, it also suggests the nature of this correspondence, i.e. FCD nuclei seem to have some (at least phonological/historical) relation to the *long* vowels.

All of the above assumptions are essentially empirically testable. If they can be shown to hold, the question is reasonable. In order for these assumptions to hold, a number of necessary conditions must hold as well. In order to show that DRESS and SQUARE are in some

sense ‘the same’, two conditions should hold: First of all, there should be some correlation between the two across the sample. In addition, their formant frequency distributions should not be too different from each other since they would be assumed to be instances of the same vowel. In order to show that KIT and the nucleus of NEAR stand in some relation to each other, it would be necessary to find similarities along a dimension other than formant frequency (as with regard to formant frequency, they would of course not match, which is implied by the question in the first place). If such a variable can be identified, we would expect to find, along that variable, the same relationship between KIT and NEAR as we do between DRESS and SQUARE. Since we are concerned with a vowel system where subsystems are distinguished by both quality and quantity, vowel duration is probably a good choice for testing these relationships. The discussion will therefore focus on both vowel quality and quantity.

It will be argued that to some extent, the implicit assumptions outlined above are all to some extent questionable with regard to the data at hand. More specifically, we will argue that if the NEAR nucleus can be likened to an element from the monophthongal subsystem at all, it would be the FLEECE vowel. Therefore, the asymmetry between the SFV chain shift and the FCD merger is not too puzzling a phenomenon, since we are not dealing with elements of the same subtype. It seems comment on why the question ‘Why did NEAR and SQUARE merge, whereas KIT and DRESS did not?’ should occur in the first place. After all, if we look at the history of English, we are dealing with vowels that originate in different subsystems. Whereas KIT as well as the majority of DRESS words stem from short vowels, the nuclei of both NEAR and SQUARE are reflexes of ME long vowels (/e:/, /ɛ:/, /ai/ and /a:/). Assuming that the general principles of vowel shifting outlined by Labov are generally correct, we would not expect vowels from different subsystems to behave in similar ways.

However, it seems to be a fact of many varieties of contemporary English that the nucleus of NEAR has a more central articulation compared to FLEECE (and is therefore qualitatively closer to KIT). This would suggest a large-scale transfer in lexical set membership at some stage in the recent history of English of formerly pre-rhotic long vowels from FLEECE (or some historical antecedent thereof) to KIT. Although this process may not have applied across the board in all varieties of English (cf. the discussion in Wells (1982: 153-54)), it was probably part of the RP standard (and represented as such in most standard dictionaries, i.e. NEAR transcribed as /ɪə/ rather than /iə/).

## 5.2 Quality – Formant frequency averages and shapes of distributions

Figure 5.1 plots the average F1/F2 values of NEAR/SQUARE/KIT/DRESS/FLEECE of the entire sample of intermediate speakers. Table 5.1 indicates Hotelling-Lawley trace scores for each possible vowel pair. Table 5.2 gives correlation test results for both formants of all the vowels under discussion.

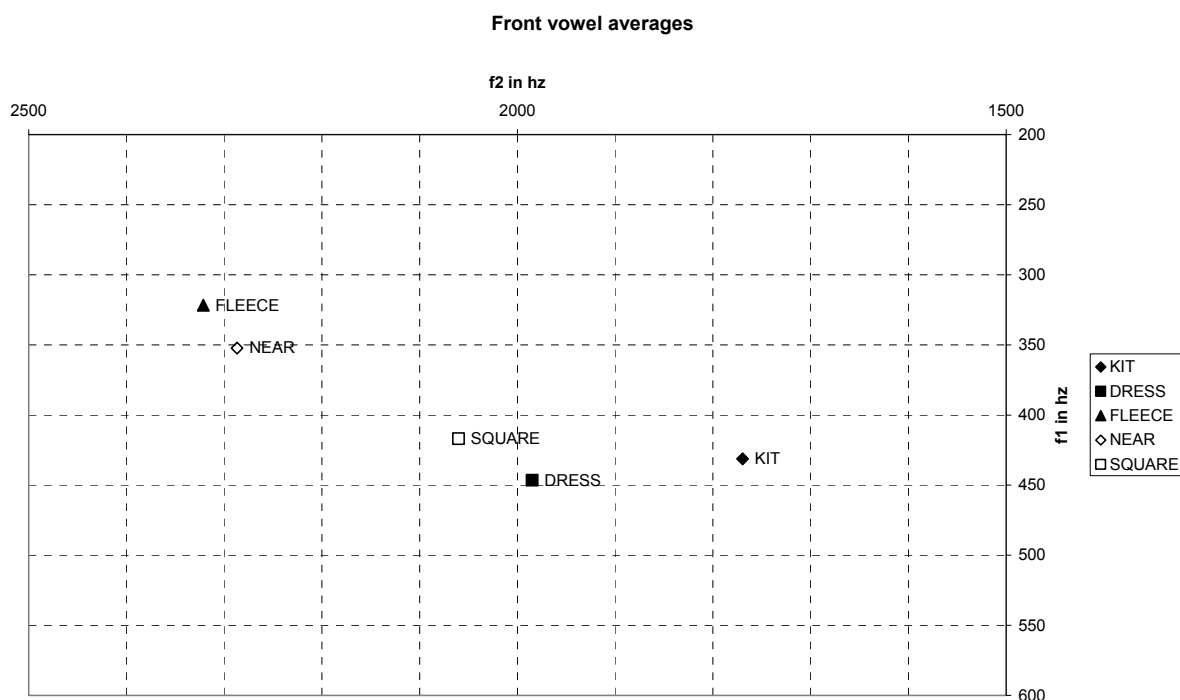


Figure 5.1 – Formant frequency averages of FLEECE – NEAR – SQUARE – DRESS – KIT. The values plotted represent the average of all speaker averages for any of the given vowels.

vowel		NEAR	SQUARE	KIT	DRESS	FLEECE
NEAR	H-L		1.872	4.27	4.37	0.1257
	p		<.0001	<.0001	<.0001	0.04
SQUARE	H-L			1.41	0.26	3.84
	p			<.0001	0.02	<.0001
KIT	H-L				1.7	6.76
	p				<.0001	<.0001
DRESS	H-L					5.02
	p					<.0001

Table 5.1 – Hotelling-Lawley trace scores and significance levels between the distributions of FLEECE – NEAR – SQUARE – DRESS – KIT. The data points on the basis of which these scores were obtained are the average formant frequency values for each speaker.

			KIT	DRESS		NEAR		SQUARE		FLEECE	
			<i>F1</i>	<i>F1</i>	<i>F2</i>	<i>F1</i>	<i>F2</i>	<i>F1</i>	<i>F2</i>	<i>F1</i>	<i>F2</i>
KIT	<i>F1</i>	rho		-0.21	<i>0.63</i>	-0.26	-0.06	0.02	0.3	0.31	<i>0.36</i>
	p			<i>0.27</i>	<i>0.09</i>	<i>0.17</i>	<i>0.74</i>	<i>0.88</i>	<i>0.11</i>	-0.19	<i>0.06</i>
	<i>F2</i>	rho	<b>-0.5</b>	<b>0.5</b>	-0.23	<b>0.46</b>	0.02	0.36	-0.02	0.12	<b>-0.52</b>
	p		<b>&lt;0.01</b>	<b>&lt;0.01</b>	0.23	<b>&lt;.05</b>	0.97	<i>0.05</i>	0.88	0.53	<b>&lt;.01</b>
DRESS	<i>F1</i>	rho			<b>-0.69</b>	<i>0.32</i>	-0.02	<b>0.53</b>	-0.19	-0.22	-0.01
	p				<b>&lt;.0001</b>	<i>0.09</i>	0.83	<b>&lt;.01</b>	0.33	0.27	0.97
	<i>F2</i>	rho				-0.035	0.02	-0.31	<b>0.46</b>	<i>0.35</i>	-0.11
	p					<i>0.85</i>	0.93	0.1	<b>&lt;.05</b>	<i>0.07</i>	0.59
NEAR	<i>F1</i>	rho					-0.25	<b>0.61</b>	-0.03	<b>0.41</b>	-0.34
	p						<i>0.19</i>	<b>&lt;.001</b>	0.86	<b>0.03</b>	<i>0.07</i>
	<i>F2</i>	rho						-0.08	<b>0.48</b>	-0.17	<b>0.47</b>
	p							<i>0.65</i>	<b>0.01</b>	0.39	<b>&lt;.05</b>
SQUARE	<i>F1</i>	rho							-0.07	0.04	0.11
	p								<i>0.71</i>	0.84	0.57
	<i>F2</i>	rho								-0.08	0.2
	p									<i>0.67</i>	0.3
FLEECE	<i>F1</i>	rho									<b>-0.51</b>
	p										<b>&lt;.01</b>

Table 5.2 – Correlation coefficients and significance levels between the individual speaker averages of FLEECE – NEAR – SQUARE – DRESS – KIT (Wilcoxon rank-sum test). Correlations that are significant below  $p = .05$  are given in bold, trends (i.e. below  $p = .1$ ) are given in italics.

Figure 5.1 shows that the nucleus of NEAR is rather close to the FLEECE nucleus. Essentially the same condition holds between SQUARE and DRESS. Although the distances between the distributions of both NEAR/SQUARE and SQUARE/DRESS are statistically significant, they are certainly smaller than those between any other two vowels. In addition, we note that the distance between NEAR and FLEECE is indeed smaller than that between SQUARE and FLEECE. Whether these differences are due to different environmental factors between the diphthongal nuclei on the one hand and the monophthongs on the other hand or whether the four distributions represent four different vowels is not entirely clear, what seems to be ruled out, however, is the assumption that SQUARE and DRESS are in some sense ‘the same’, whereas NEAR and FLEECE are not. Under this assumption, we would expect the distance between NEAR and FLEECE to be at least as large (but probably larger) than that between SQUARE and DRESS.

If we look at the correlations given in table 5.2, we see that both NEAR and FLEECE as well as DRESS and SQUARE show positive correlations with each other for both formants. That

is, the higher/fronter the NEAR nucleus, the higher/fronter the FLEECE vowel. The same relation holds between SQUARE and DRESS. In addition, both centring diphthongs correlate positively with each other, which is not the case between NEAR and DRESS. This essentially means that adjacent segments correlate, but not others, which rules out a ‘static vowel space’ approach whereby the correlations between two vowels are explained solely on the basis of individual systems, since in this case, we would probably expect correlations between all vowels. Rather, what we seem to be observing here is a dynamic development in the vowel system which implies not only the SFV shift as described in chapter 3, but also the front centring diphthongs as well as the FLEECE vowel. What is less clear, however, is whether the correlations between e.g. DRESS and SQUARE reflect a functional relationship between the two vowels or merely different degrees of innovativeness within a sample. That is, does the SQUARE nucleus correlate with DRESS because it is the same vowel, or does DRESS push out SQUARE, or do they simply rise independently of each other over the same period of time? If they were the same vowel, we would probably expect a somewhat tighter correlation; in addition, we would expect the correlation to hold approximately equally across the entire sample. Figure 5.2 plots the first formant frequency averages of DRESS and SQUARE from each intermediate speaker against each other. The smoothed correlation line is roughly parabolic, and becomes approximately parallel to the x-axis toward high F1 values on the x-axis (i.e. high DRESS F1s), which suggests that the more conservative a given speaker’s vowel system, the weaker the correlation between SQUARE and DRESS in the F1 dimension.

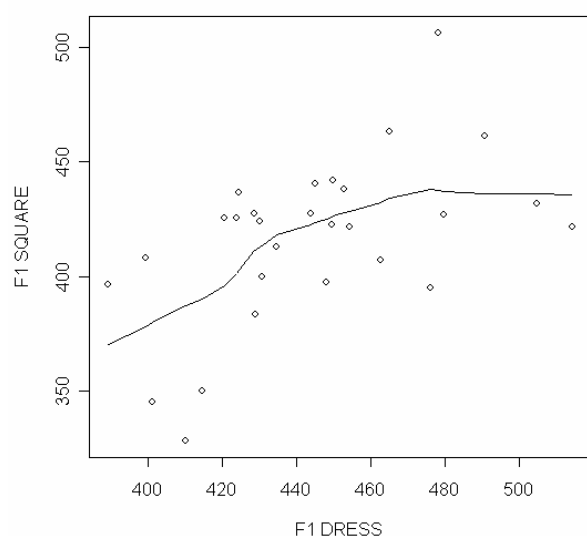


Figure 5.2 – Individual mean F1 values of DRESS (x-axis) and SQUARE (y-axis). The line represents a non-parametric scatterplot smoother fit through the data (Cleveland 1979).

This allows for two possible conclusions: Either the two vowels move independently of each other, and the onset of DRESS raising occurs before SQUARE starts moving upwards, or the nucleus of SQUARE becomes associated with the DRESS vowel as DRESS moves upwards and approaches SQUARE. The two scenarios are sketched in figure 5.3.

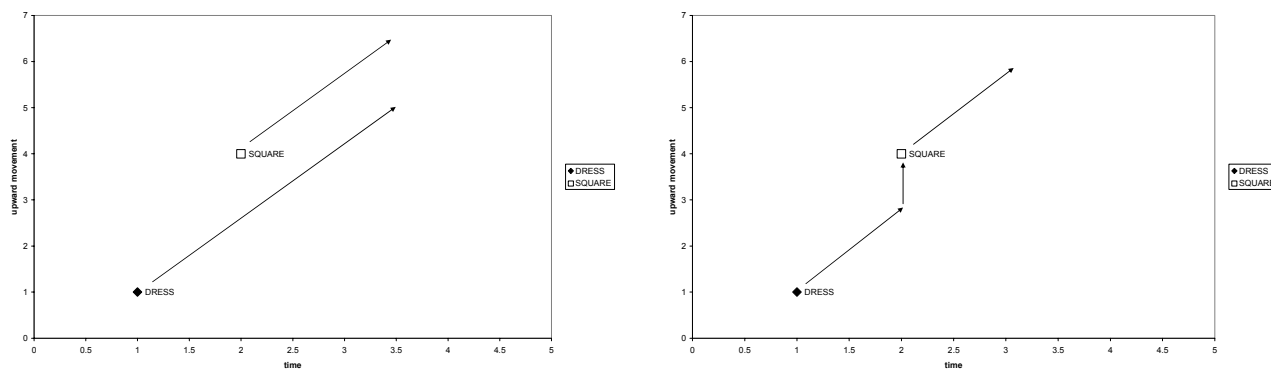


Figure 5.3 - Two potential scenarios of DRESS and SQUARE raising under increasing correlation over time: Independent raising of the two nuclei (fig. 5.3 (a)) and association of DRESS with SQUARE during the process (5.3 (b)).

Under the first scenario (fig. 5.3 (a)), we would probably expect statistically significant differences between the two distributions throughout the entire sample, whereas under the ‘merger’-scenario, we would expect the two distributions to merge along the way. The evidence, however, is somewhat contradictory: If only speaker averages are taken into account, it is only in the early and medium sample where the nuclei of SQUARE and DRESS retain significantly different distributions. If the entire data pool is taken into account, the two distributions remain significant for all age groups. Table 5.3 shows Hotelling-Lawley trace scores for both data pools.

Averages			
	early	medium	late
H-L trace	0.44	0.43	0.31
p	<.05	<.05	0.16

Table 5.3 (a)

<i>All datapoints</i>			
	early	medium	late
H-L trace	0.09	0.026	0.03
p	<.0001	<.0001	<.001

Table 5.3 (b)

Table 5.3 – Hotelling-Lawley trace scores for distributions of DRESS and SQUARE. Table 4.9 (a) shows the values for distributions consisting of speaker averages, 4.9 (b) shows all data points.

age	sex	speaker	H-L trace	p
early	female	<b>1</b>	<b>0.17</b>	<b>&lt;.01</b>
		<b>2</b>	<b>0.2</b>	<b>&lt;.01</b>
		3	0.03	>.2
		4	0.03	>.2
		5	0.07	>.7
		<b>6</b>	<b>0.25</b>	<b>&lt;.0001</b>
	male	<b>1</b>	<b>0.68</b>	<b>&lt;.001</b>
		2	0.03	>.1
		<b>3</b>	<b>0.29</b>	<b>&lt;.001</b>
		<b>4</b>	<b>0.49</b>	<b>&lt;.0001</b>
medium	female	1	0.02	>.4
		2	0.2	>.4
		3	0.07	>.05
		<b>4</b>	<b>0.19</b>	<b>&lt;.001</b>
		5	0.019	>.4
	male	1	0.01	>.6
		2	0	>.9
		<b>3</b>	<b>0.23</b>	<b>&lt;.001</b>
		4	0.09	>.6
		<b>5</b>	<b>0.19</b>	<b>&lt;.0001</b>
late	female	<b>1</b>	<b>0.29</b>	<b>&lt;.0001</b>
		2	0.02	>.5
		<b>3</b>	<b>0.13</b>	<b>&lt;.01</b>
		4	0.1	>.05
	male	<b>1</b>	<b>0.2</b>	<b>&lt;.01</b>
		2	0.01	>.6
		3	0.03	>.2
		<b>4</b>	<b>0.1</b>	<b>&lt;.05</b>

Table 5.4 – Hotelling-Lawley trace scores for DRESS and SQUARE for each individual Intermediate speaker.

It seems clear, however, that if we compare the early and the late speaker groups, the distance between the two distributions becomes smaller. As for the medium group, the evidence is less clear. If we look at speaker averages (table 5.3 (a)), the medium group is rather more similar to the early group, in that they both retain significantly different distributions of DRESS and



SQUARE. In addition, the distance between the two is very similar. If we look at the pooled data points (table 5.3 (b)), we can see that the distance between the two vowels in the medium sample is more similar to (in fact, even smaller than) the late group.

It might therefore be advisable to look at the two vowels for each individual speaker (table 5.4). Given the data in table 5.4, no clear conclusion can be drawn as to whether the relation between the SQUARE nucleus and the DRESS vowel over the Intermediate period, since there are speakers with both distinct as well as non-distinct distributions within each cell. We may therefore have to conclude that either one of the putative scenarios depicted in figure three may be the case. As for the difference between NEAR and FLEECE, the case seems to be somewhat clearer.

<i>Averages</i>			
	early	medium	late
H-L trace	0.46	0.62	0.01
p	<.05	<.05	>.9

Table 5.5 (a)

<i>All datapoints</i>			
	early	medium	late
H-L trace	1	0.54	0.14
p	<.0001	<.0001	<.0001

Table 5.5 (b)

Table 5.5 – Hotelling-Lawley trace scores for distributions of NEAR and FLEECE. Table 5.5 (a) shows the values for distributions consisting of speaker averages, 5.5 (b) shows all data points.

If we look at speaker averages (table 5.5(a)), it seems as if the distributions of NEAR and FLEECE were significantly different for the early and the medium group only. However, the pooled data (table 5.5(b)) suggest a significant difference between the two vowels in all speaker groups. In addition, all individual speakers except one (late male) keep the two

vowels distinct. We do, however, note an appreciable decrease in the overall distance between NEAR and FLEECE in both data sets. Figure 5.4 plots all tokens of FLEECE/NEAR for the early and the late sample.

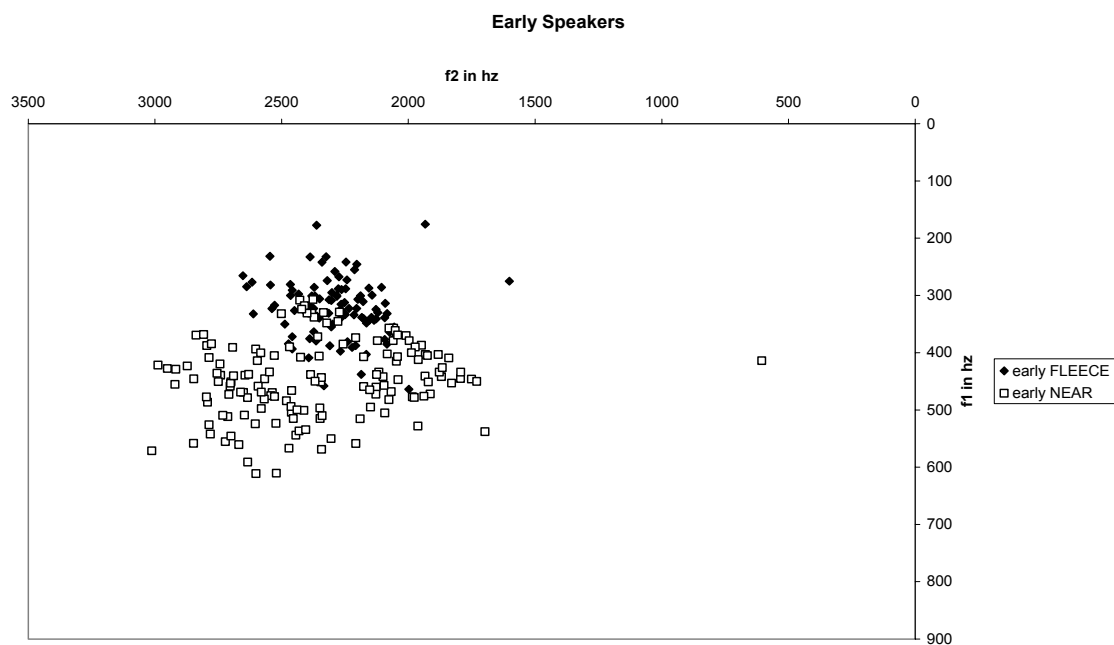


Figure 5.4 (a)

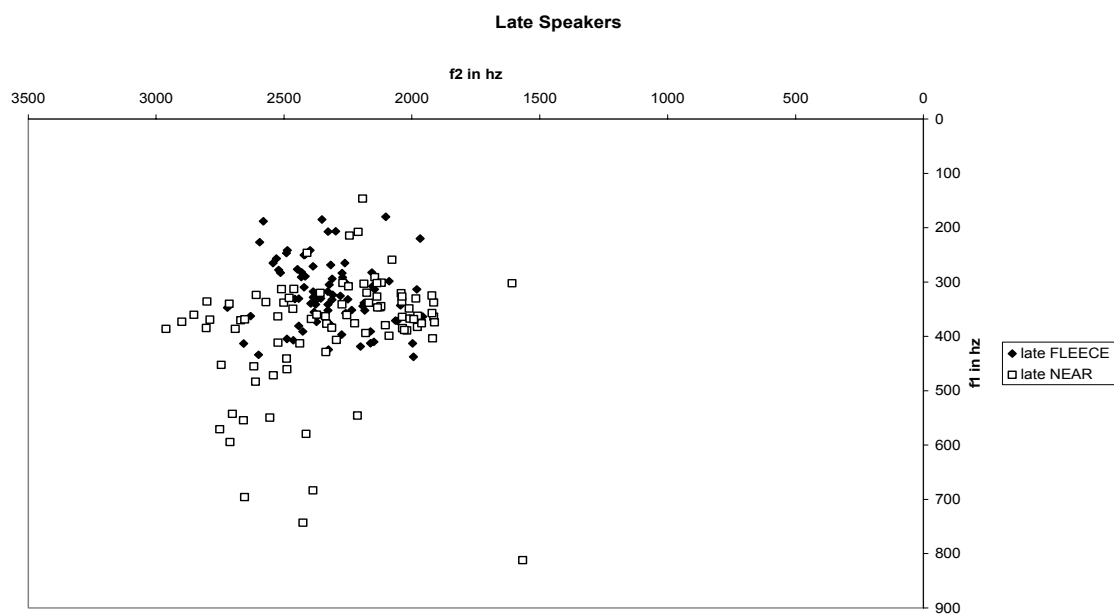


Figure 5.4 (b)

Figure 5.4 – Formant frequency values of NEAR and FLEECE in the early (5.4 (a)) and the late (5.4 (b)) sample.

It seems clear that the overall distribution of NEAR shifts upwards toward the vowel space occupied by FLEECE over the Intermediate period. In addition, what we noted in chapter 3 above with respect to KIT seems to hold for NEAR as well, namely that the space available for a given vowel expands along the trajectory of the shift. A further property of NEAR in the current sample is the complete absence of central variants at any stage of the developments analysed here. We therefore have further evidence that the NEAR nucleus is (whatever else it might be) not a KIT vowel in the Intermediate period. Recall from the data given in table 5.2 above that the first formant frequency of NEAR and the second formant frequency of KIT are correlated to the effect that the lower the F1 of NEAR, the lower the F2 of KIT, whereas the same formants for each of the two vowels show no such correlation. This shows that over the period analysed in the present study, the two vowels occupy different areas in articulatory space and shift along different pathways, which suggests that they are indeed different. The question here is to what extent we can extrapolate backwards: After all, given that KIT centralises, it might well be the case that for the most conservative speakers in the sample, KIT and NEAR may show a rather close connection. This, however, does not seem to be the case. Figure 5.5 plots all tokens of NEAR, FLEECE and KIT for the early male speakers; table 5.6 shows Hotelling-Lawley trace scores between the three distributions.

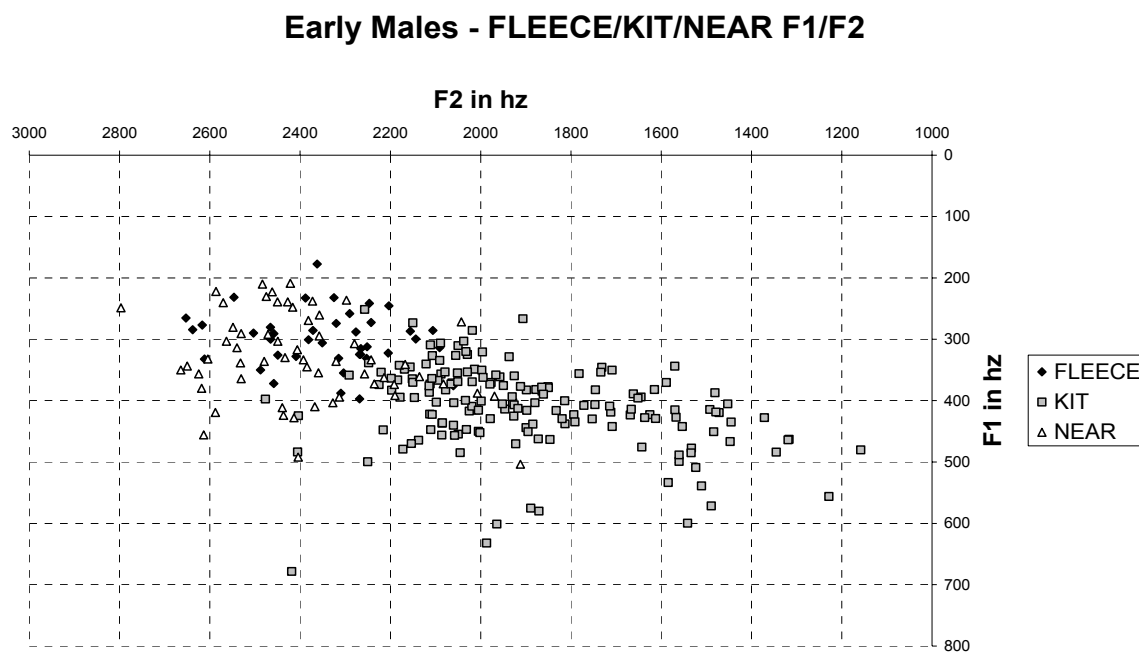


Figure 5.5 – F1/F2 plot of all tokens of NEAR, FLEECE and KIT in the early male sample.

	NEAR		KIT	
	<i>H-L trace</i>	<i>p</i>	<i>H-L trace</i>	<i>p</i>
<b>FLEECE</b>	0.1	<.01	.79	<.0001
<b>NEAR</b>			.89	<.0001

Table 5.6 – Hotelling-Lawly trace scores and significance levels between FLEECE, NEAR and KIT in the early male sample.

Figure 5.5 shows that even for the most conservative speaker group, the overlap between NEAR and FLEECE is considerable, whereas KIT is rather well separated from the two. This is reflected by the distance measures given in table 5.6: Although the differences between all three distributions are statistically significant, the distance between KIT and both NEAR and FLEECE is appreciably higher than that between NEAR and FLEECE. It therefore seems that for the Intermediate period, there is little indication of a ‘more KIT-like near becoming less KIT-like’ sort of process. If there ever was an association (and subsequent dissociation) of the NEAR nucleus with the KIT vowel, it must have taken place earlier. From the formant frequency analysis, we can therefore conclude that

- (1) The status of the SQUARE nucleus is not clear. Whereas some speakers show significantly different distributions of SQUARE and DRESS, others do not. Although the pooled data indicates that the distance between the nuclei decreases over the Intermediate period, this trend is not reflected in the data from individual speakers.
- (2) Almost all speakers keep their distributions of NEAR and FLEECE distinct, although the overall distance between the two decreases over the sample.
- (3) The nuclei of NEAR and KIT are distinct over the entire sample. There are, at no time, any appreciable numbers of NEAR tokens which are more central than the most central FLEECE tokens.

- (4) For the questions outlined in section 5.1, this implies that the assumption that the nuclei of the centring diphthongs can generally be associated with a vowel from the monophthong system is not fully borne out. Rather, there is some indication that the NEAR nucleus is different from both KIT and FLEECE, although the distance between NEAR and FLEECE decreases over the Intermediate period.
- (5) If anything, there is rather more support to the assumption that ‘SQUARE is to DRESS as NEAR is to FLEECE’ since the members of each pair correlate with the other one in the F1 dimension.

### 5.3 Duration revisited

In addition to the formant frequency data discussed above, it might be helpful to consider the duration dimension in assessing the relation between the diphthongal nuclei and the monophthongs. As we have seen in chapter 3 above, there are solid durational differences between KIT and DRESS across all speaker groups whereby the former vowel is about 50% the duration of the latter in monosyllables (~50 ms for KIT vs. ~100 ms for DRESS). If the diphthongal nuclei were the same as the short front monophthongs, we would probably expect this difference to be reflected in the duration of the centring diphthongs as well. This, however, does not seem to be the case. The overall duration of NEAR is 175.7 ms compared to 164.8 ms for SQUARE in monosyllables (cf. table 4.7). If we assume identity of the diphthongal nuclei with the short monophthongs, we would also have to assume radically different nucleus-to-offglide proportions in each of the two front centring diphthongs. If we compare only the overall averages, we would calculate a offglide/nucleus duration ratio of  $126/50 = 2.52$  for NEAR and  $64.8/100 = 0.65$  for SQUARE. This seems unlikely, but may not be altogether impossible. It is notoriously difficult to impose a binary segmentation (nucleus+offglide) upon what is essentially an articulatory continuum (motion of the articulator from one position to some other position), and any such segmentation can probably be challenged. However, if it could be shown that a nucleus/offglide boundary assigned to the overall trajectory on the basis of the duration averages (as calculated on the

basis of monophthong durations) subdivides the two centring diphthongs at radically different points in their overall formant trajectory, this would probably be indicative of a non-identity of diphthongal nuclei and the short monophthongs in at least one of the two cases. Figure 5.6 below shows formant trajectories of both centring diphthongs as scatterplot smoothed lines through all data points. The vertical lines indicate the nucleus/offglide boundary under the assumption that the duration of KIT and DRESS equal that of the nucleus of NEAR and SQUARE, respectively.

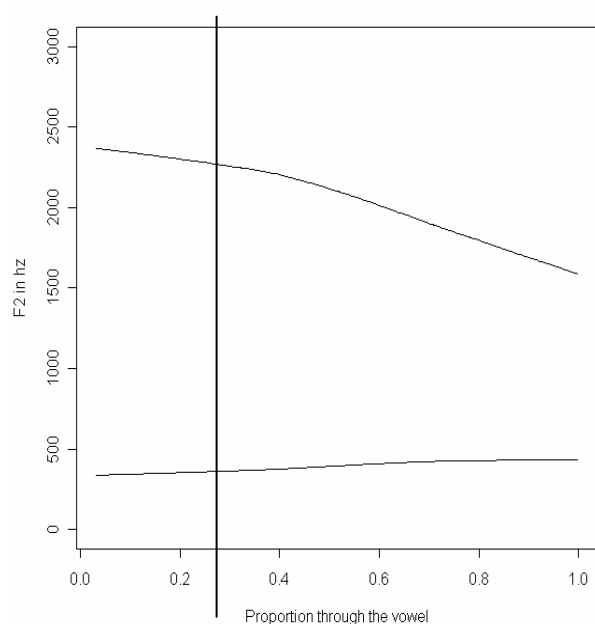


Figure 5.6 (a)

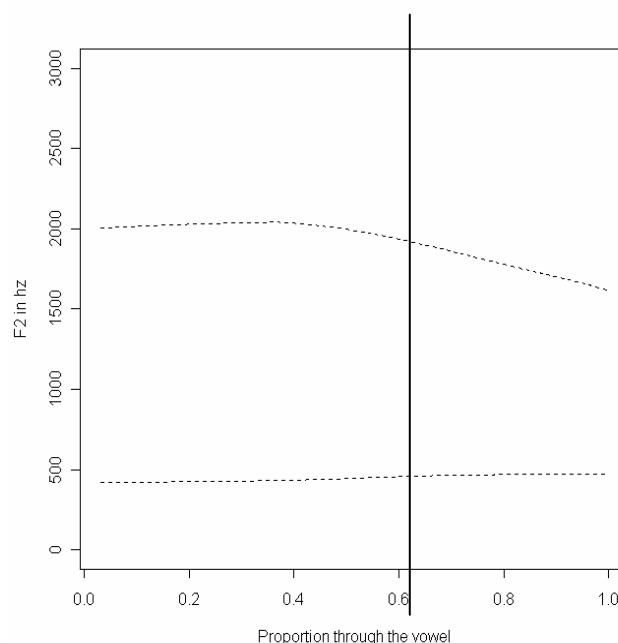


Figure 5.6 (b)

Figure 5.6 – Average time-normalised formant frequency trajectories of NEAR (5.6 (a)) and SQUARE (5.6 (b)) in the Intermediate sample. The space between the onset of the formant movement and the vertical line indicates the approximate duration of KIT (5.6 (a)) and DRESS (5.6 (b)) relative to the centring diphthongs.

It seems clear that if we look at overall duration of KIT/DRESS and extrapolate from there to the nucleus/offglide ratio in the centring diphthongs, the internal segmentation along the duration axis is quite different for each of the two diphthongs: Whereas for NEAR, the duration of the nucleus is considerably shorter than that of the offglide, the inverse situation holds for SQUARE. However, since we are concerned with different starting points and a

similar endpoint for each of the two diphthongs, we may want to apply another measure, namely the extent to which a diphthong trajectory has reached its target at the point of the boundary as calculated above. This can be expressed as the ratio of a formant value at some first prominent point throughout the trajectory (nucleus) and some prominent second point (offglide). Figure 5.7 illustrates this: The overall distance covered in formant space (here F2) is equivalent to  $a$ . The ratio  $b/a$  expresses the extent to which  $a$  has been traversed at the boundary. The higher the ratio, the further through the overall trajectory the boundary.

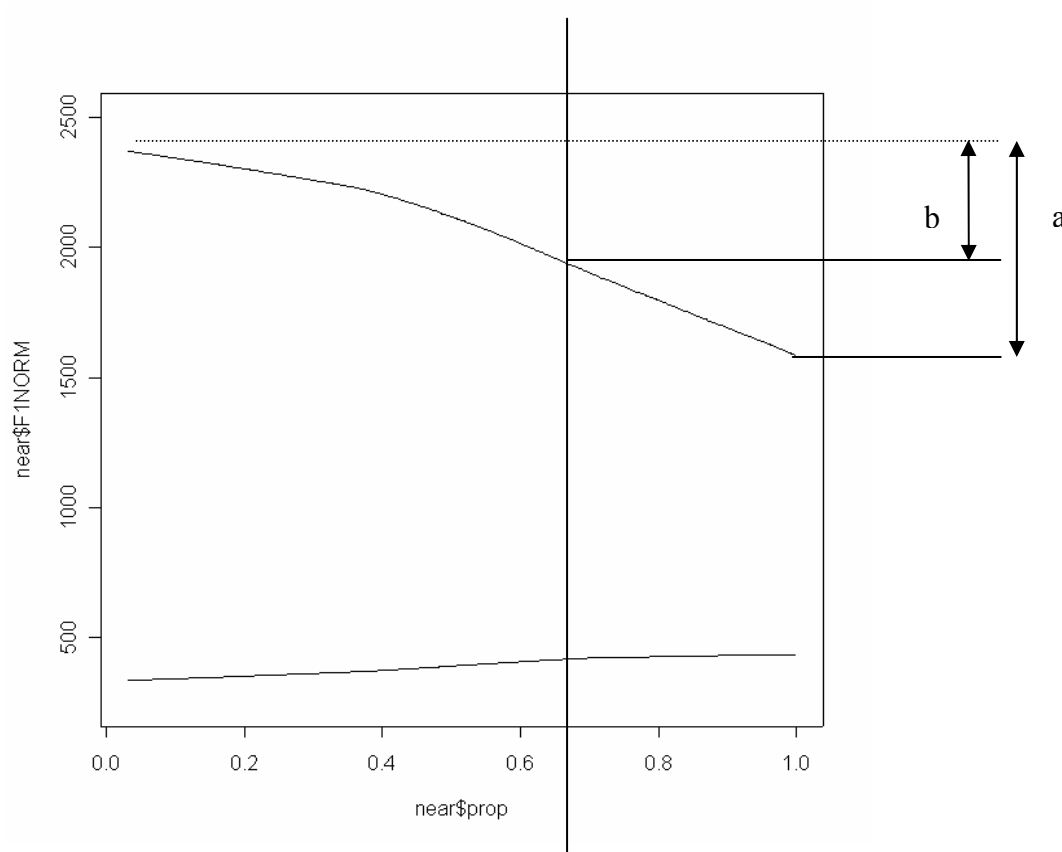


Figure 5.7 – Proportion-through-trajectory measure, schematised:  $b$  represents the frequency range between the highest point in the trajectory and the point that correlates with the average duration of some reference monophthong.  $a$  represents the entire frequency range of a given formant over the entire trajectory.

The reference numbers have been defined as follows:

First prominent 'point' = average  $F_n$  value between .0 and .2 throughout the diphthong.

Second prominent point = average values between .8 and 1 throughout the diphthong.

In addition, the boundaries as given by the KIT and DRESS duration measures (.28 in NEAR and .6 in SQUARE) have been redefined dynamically in order to minimise random skew due to small numbers of observations at the specific points. In both cases, the average  $F_n$  within the range  $\pm .03$  were calculated. Therefore, the formant frequency value for the boundary marked by the KIT duration in NEAR is the average  $F_n$  of all observations .25 - .31 throughout the overall trajectory.

If we apply this ratio to the KIT average duration relative to the NEAR trajectory (as in fig. 5 (a)), we obtain the following values:  $b/a(F1) = .08$ ,  $b/a(F2) = .05$ . For SQUARE (cf. 5 (b)), we obtain  $b/a(F1) = .61$  and  $b/a(F2) = .48$ . This means that by the time the NEAR diphthong has reached the duration of an average KIT vowel in the present sample, it has traversed  $\sim 8\%$  of its overall  $F1$  range (and  $5\%$  of its  $F2$ ). In the case of SQUARE, the formant trajectory movement has proceeded much further as the diphthong reaches the length of DRESS ( $\sim 61\%$  for DRESS and  $\sim 48\%$  for SQUARE).

It therefore seems clear that a proportion-through-trajectory measure yields further evidence for the proposition that NEAR and KIT stand in a different relation to each other than do SQUARE and DRESS<sup>35</sup>.

#### 5.4 Conclusion

The data analysed here does not allow for a satisfying answer to the question 'what is a front centring diphthong nucleus'. We have seen in section 5.2 that the distribution of the nucleus of SQUARE is different from that of DRESS for some speakers, but not for others and there is no clear directional trend in favour of either one of the two possibilities. There is some

---

<sup>35</sup> It would of course have been desirable to obtain similar calculations on the basis of the duration properties of the FLEECE vowel. However, no duration measurements of FLEECE were taken within the course of the research reported here.



indication that the NEAR nucleus approaches the FLEECE vowel, but on the other hand, the two distributions are distinct for all speakers except one. We might have to consider the diphthongal nuclei as different entities which do not necessarily correspond to anything in the monophthongal system altogether. Given this, the inferences from the monophthong system to the diphthong system as outlined in the introductory section are to some extent invalidated. We have seen that the assumption that ‘NEAR is to KIT as SQUARE is to DRESS’ does not hold up to the data analysed here. This conclusion is backed by the discussion of both formant frequency as well as duration. It therefore seems that the question ‘Why did NEAR/SQUARE merge, whereas DRESS pushed out KIT’ seems somewhat less puzzling if we take into account the lack of equivalence between the SFVs and the centring diphthongs.

Even if we regard the diphthongal nuclei as being equivalent to the short front vowels, we could regard the different trajectories of change (i.e. merger vs. chain shift) as an instance of ‘swamping’ of distinguishing cues in an intermediate stage: I have argued in chapter 3 that even for speakers who show overlap between KIT and DRESS in F1/F2 space, the two vowels remain distinct if we factor in duration. In the case of the FCDs, however, this is not the case, since the offglide (which remains invariant across the two vowels) contributes to the overall duration of the diphthongs, which results in a less solid duration difference between the two.

## 6.

## Broad A in Intermediate NZE

## 6.1 Introduction

In the following chapter I will discuss a further variable that, to some extent at least, sets off contemporary NZE from AusE. This is the so-called DANCE/BATH vowel. Section 6.2 gives the necessary terminology and provides a brief survey of the current dialect situation with respect to that vowel as well as summing up the relevant historical developments. Section 6.3 sums up findings from MU data. Data from the Intermediate period is discussed in 6.4 – 6.7. Section 6.4 shows formant frequency data and provides a CART analysis of relevant conditioning factors. Some implications of this will be discussed in 6.5 with regard to the mechanisms of the merger. 6.6 presents data from duration measurements. Finally, the role of proper nouns will be discussed in section 6.7

## 6.2 Background

The ‘d[a:]nce - d[æ]nce’ distinction, which now sets off Southern British English from most North American varieties as well as New Zealand English from Australian English is the last stage of a rather long and complex chain of developments in the English low vowel system, of which I will here provide a brief summary. This alternation is also referred to as the ‘broad A rule’, where ‘broad A’ refers to the long open vowel. Note that most of the following outline is based on the more comprehensive discussion in Lass (1999).

During the early seventeenth century, a number of Southern varieties of English (whence all the contemporary ‘Standard’ extraterritorial dialects, cf. Lass (1990)) fronted the short ME /a/ nucleus to a value between Cardinal 3 and 4, i.e. [æ]. This left Early Modern English bereft of a low non-rounded vowel, and much of the later history up to the merger of transfer between TRAP and START can be regarded as a filling up of that gap. Later on in the seventeenth century, the recently created /æ/ sound seems to have been lengthened to [æ:] in some environments (especially before /r/, cf. ‘far’, as well as frequently before fricatives and

nasals if followed by another consonant, cf. ‘dance’ and ‘castle’), but not in others. The latter group consists of tokens that are now subsumed under the lexical set of TRAP. From the lengthening process, a new phoneme is born (START).

Note that Lass’ analysis posits a lengthening process of /æ/ independently of whether the word in question contains a rhotic element in its coda. Therefore, under this analysis, the first stage of the TRAP – START split is an allophonisation that splits up all items of the two current lexical sets, rather than a process of loss of non-prevocalic ‘r’ followed by compensatory lengthening of the vowel (‘far’, ‘car’) and a subsequent transfer of other tokens to that class. This implicitly makes a prediction with regard to the data we will analyse below, namely that TRAP items that eventually move over to the START class should ideally be longer than the ones that do not while retaining the front nucleus. We will discuss the evidence for and against this implication below.

The next stage of the process involved the backing of the lengthened nucleus to [a:]. In varieties which underwent the lengthening process in the first place, backing is obligatory before etymological ‘r’, whether still realised or not (as in ‘far’ or ‘car’, which have a non-front nucleus in all varieties in question at some stage in the process). It is variable in other environments. This variability defines some salient isoglosses and can be simplified as follows: In pre-fricative lengthening environments (‘chaff’, ‘castle’), Southern British English as well as the Southern Hemisphere Englishes have the non-front nucleus. Most North American varieties outside New England have retained a front nucleus. Before nasals, a similar situation holds with the notable exception of Australia, where both types of nuclei are commonly found (cf. Wells 1982). With regard to the above mentioned varieties, the situation can be summed up as follows on the basis of the previous discussion (cf. Lass 1999, p. 104):

	RP	Northern British	GenAm	Australia	NZ
TRAP	[æ]	[a]	[æ]	[æ]	[æ]
BATH	[a:]	[a]	[æ:]	[a:]	[a:]
DANCE	[a:]	[a]	[æ:]	[æ] ~ [a:]	[a:]
START	[a:]	[a:]	[a:]	[a:]	[a:] <sup>36</sup>

Table 6.1 – Present-day reflexes of ME /a/ in different varieties of English. The lexical set names BATH and DANCE refer to specific environments where alternations between the two vowels are found, where DANCE refers to pre-nasals (such as *example, prance, slander*), and BATH refers to pre-fricatives (e.g. *path, castle, pass*).

Again, note that the lengthenings indicated in table 6.1 above stem from two different processes: In GenAm, the sole source of differential length is the ‘pure’ lengthening process as outlined above. In non-rhotic Northern British English, the lengthening difference is exclusively due to compensatory lengthening in formerly pre-rhotic nuclei. In AusE, NZE and RP it seems that the long nuclei are a product of either the lengthening process alone (DANCE, BATH) or a combination of the two (START)<sup>37</sup>.

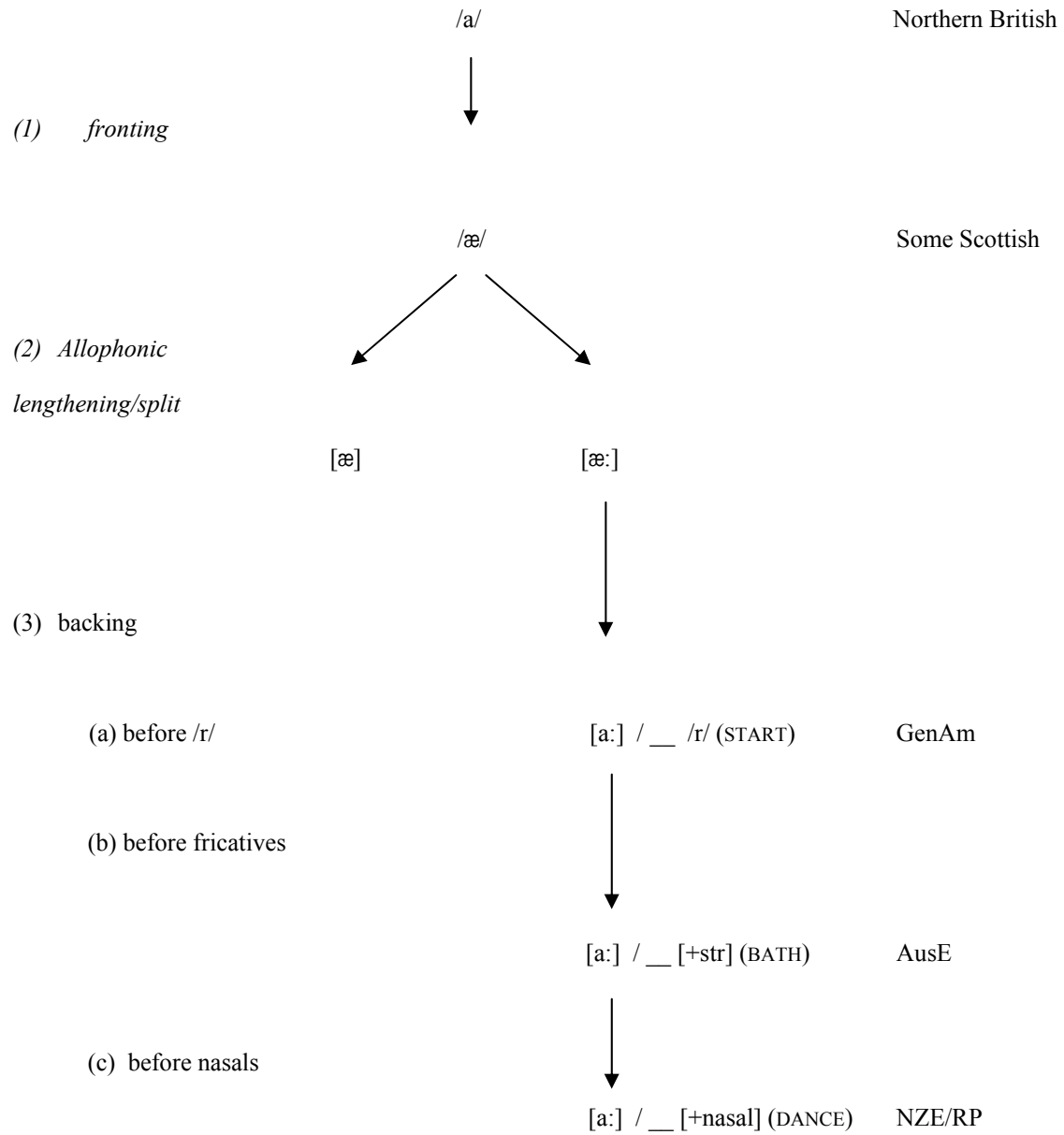
In addition to this, Lass (p. 105) notes a later process of lengthening within the newly created lexical set of TRAP, whereby vowels before nasals as well as voiced stops are distinctively longer than their counterparts before voiceless stops. This, however, seems to be a general property of many Englishes across their entire vowel space (cf. the discussion (and references) of vowel length in chapter 3). If these assumptions are correct, we would expect a number of properties in a transitional system. Note that the following line of reasoning works only under the hypothesis that the cross-dialectal patterns sketched in table 6.1 reflect a trajectory of innovation from Northern English (which never underwent the fronting process of ME /a/) to earlier Southern BE, then to GenAm (fronted and lengthened DANCE and BATH, re-backed START), to AusE (fronted and lengthened DANCE, BATH and START, re-backed BATH, partly re-backed DANCE) and finally to NZE/RP (fronted, lengthened and re-backed DANCE, BATH and START), which can be diagrammed as shown below:

<sup>36</sup> Bartlett (p.c.) notes the presence of flat A in contemporary Southland NZE.

<sup>37</sup> Although the present thesis does not address this question, it would be interesting to analyse those latter varieties as to whether there is a three-way difference in vowel length between words like START, BATH and TRAP, or whether compensatory lengthening has applied ‘vacuously’ in these varieties after the loss of non-prevocalic /r/.

process

reflected in current



This would lead to the following expectations:

- (1) In a period of transition between the GenAm stage and the RP stage, we would expect to find non-backed DANCE/BATH nuclei that are noticeably longer than TRAP nuclei (since lengthening and backing occur independently of each other, and the former precedes the latter).
- (2) Given the situation in Australia today, we would expect backing of BATH to precede the backing of DANCE.

One further point deserves mentioning in this context: Up to this point, nothing in the exposition questioned the process as an example of neo-grammarian sound change. It should be pointed out, however, that at least with respect to what we know from the current situation in Australia, NZ and Southern England, patterns of lexical diffusion seem to emerge rather obviously. That is, it is impossible to fully define phonemic environments for the change from TRAP to START. Although the above mentioned factors ‘following nasal/fricative + another consonant’ covers a considerable amount of words that went over from TRAP to START, words like ‘cancer’, ‘maths’ and ‘pass’ are obvious exceptions. In addition, there exist a small number of clear minimal pairs (cf. ‘ant’ and ‘aunt’<sup>38</sup>) synchronically, which point toward the process as a product of lexical diffusion rather than environmentally conditioned sound change.

### 6.3 The START vowel in early NZE

Gordon et al. (2004, p.125-136) point out that in their Mobile Unit data the START vowel generally had a central (low) quality, i.e. the British realisation with a back vowel [ɑ] was uncommon even for the first two generations of NZE speakers. In addition, vowel seems to have undergone fronting over the period which the Mobile Unit analysis covers.

---

<sup>38</sup> The minimal pair *ant* vs. *aunt* works only from a synchronic point of view, since *aunt* derives from ME /au/ rather than short /a/. A ‘historically proper’ near-minimal pair would be *math* (with flat A) vs. *aftermath* (broad A).

As far as lexical set membership of DANCE and BATH is concerned, it seems clear that the TRAP vowel was commonly retained in DANCE words in the Mobile Unit, where 49% of all tokens were realised with the front vowel. In addition, and contrary to the results presented here, Gordon et al. found that the TRAP realisation of DANCE was predominantly found in the speech of female speakers and remains relatively stable over that period, whereas in the male sample there occurs a marked shift toward the low vowel (p. 133). It was further pointed out that the TRAP variants were more common in coronal environments. The question of whether the alternations are due to lexical transfer or a stepwise merger-by-approximation type of change was not addressed in that study (although the methodology, i.e. binary auditory classification, implicitly assumes that the former is the case).

As for the BATH vowel, the front variant was found in only 6.6% of all tokens analysed, which suggests that historically, BATH went over to START much earlier than DANCE, which is in keeping with the data from present-day Australia (where all BATH vowels are realised with the START vowel, whereas DANCE is variable between START and lengthened TRAP).

#### 6.4 Formant Frequency distribution of the START vowel in the Intermediate period

Figure 6.1 shows formant frequency plots of all DANCE and BATH tokens obtained from the present sample in relation to the formant frequency of START and TRAP. Figure 6.1 (a) shows the position of all DANCE and START tokens in the sample; figures 6.1 (a) – (d) break up the data into the following categories: DANCE vs. BATH (fig. 6.1(b)), MALE vs. FEMALE (fig. 6.1 (c)), and EARLY vs. LATE (6.1 (d)). Also shown are the baseline vowels, which represent the formant frequency averages of all TRAP/START tokens in the present sample. Not included in the baselines are DANCE/BATH measurements, i.e. potential broad-A words.

Overall, the following observations can be made:

- (1) Most of the data points are concentrated around the average of the baseline START vowel (fig. 6.1 (a)). This shows that most speakers in the Intermediate period tend to

have the modern set-up, or something rather close to it. On the other hand, there is a considerable minority of tokens which are located in the vicinity of baseline TRAP, which suggests that, if there was a process of transfer, it was not yet fully completed for at least some of these speakers.

- (2) That this is a directional process can be inferred from the pattern in fig. 6.1 (d), which plots the data from the early and late speaker groups only. It shows that virtually all tokens that are closer to the baseline TRAP mean come from the early group, whereas the late group shows no tokens in that region, from which we can infer that at that stage, the modern set-up of BATH and DANCE vowels had been reached (i.e. they were all START vowels).

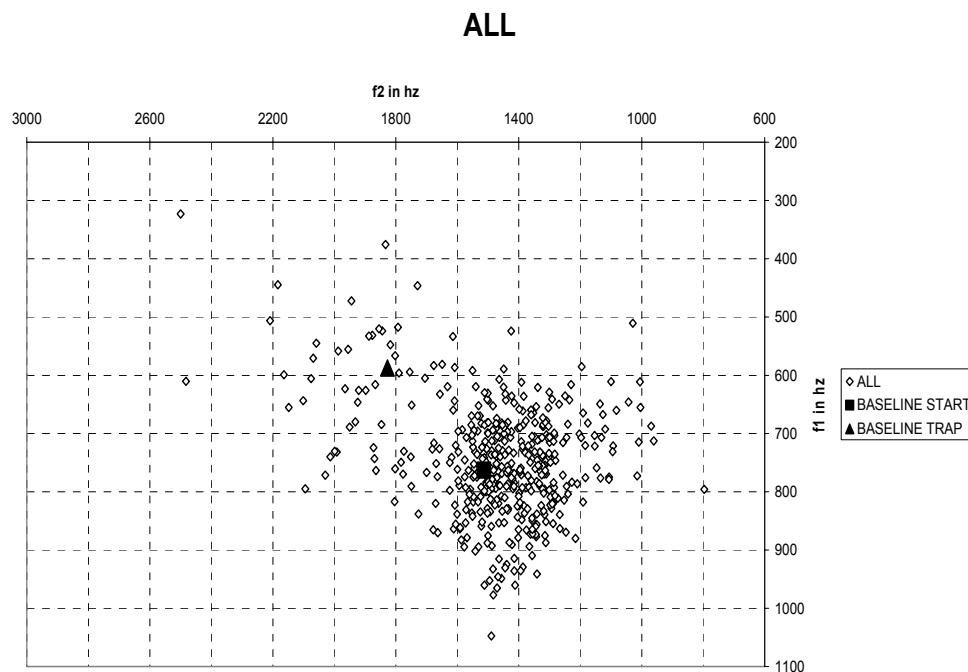


fig. 6.1 (a)



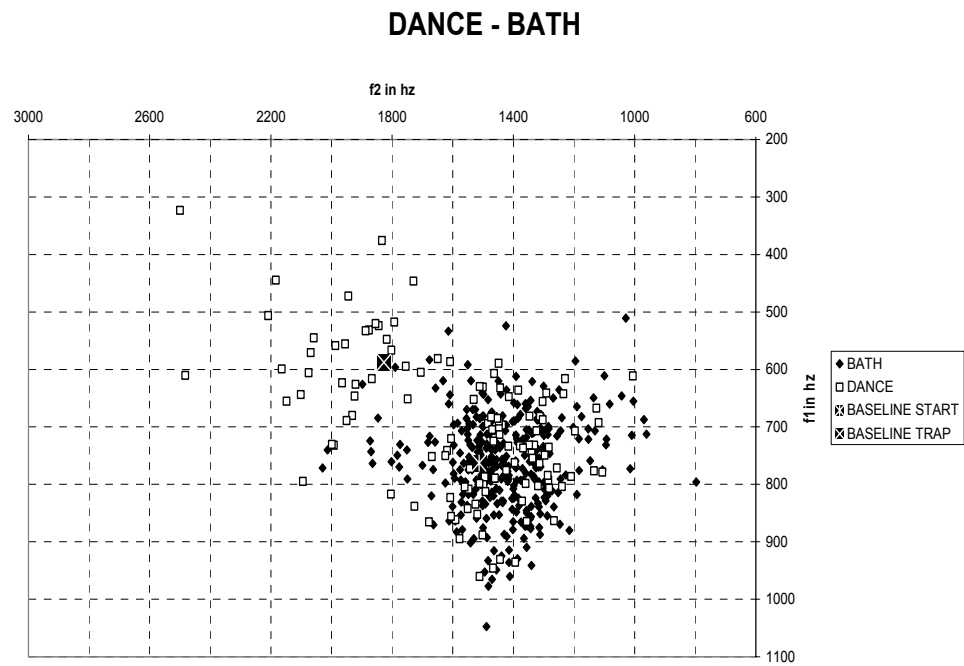


fig. 6.1 (b)

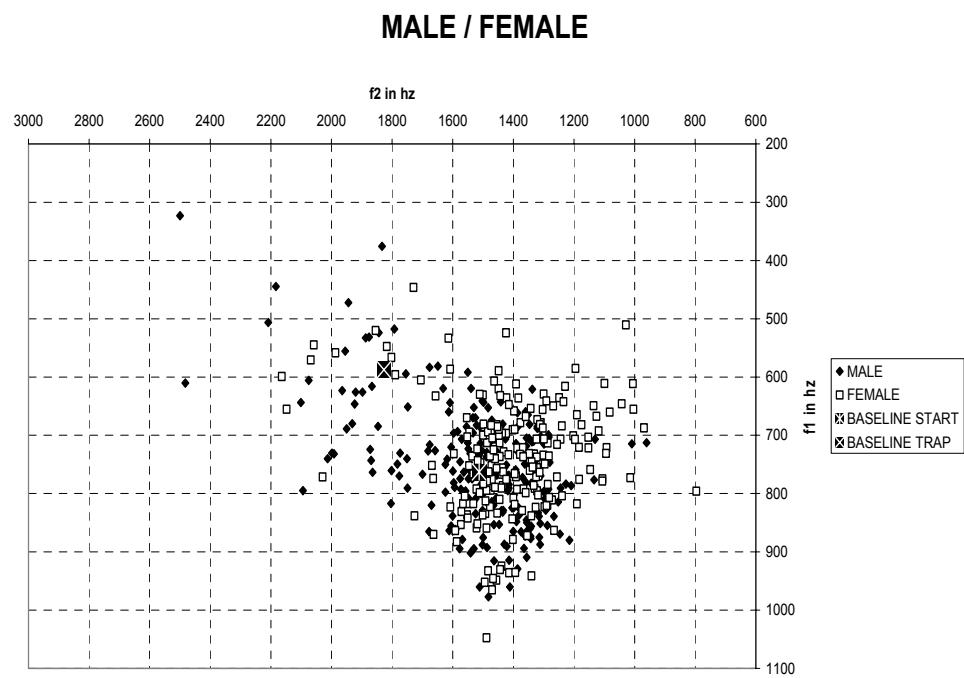


fig. 6.1 (c)

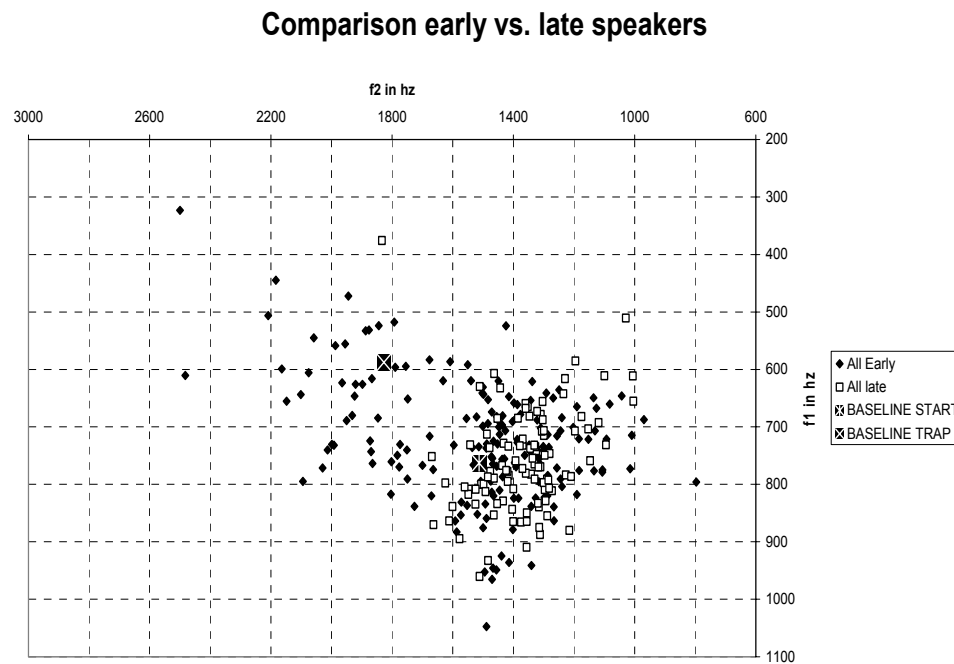


fig. 6.1 (d)

Figure 6.1 – Formant frequency plots of DANCE/ BATH against TRAP/START averages. 6.1 (a) shows all tokens from the current sample. 6.1 (b) – (d) break up the data into pre-nasal vs. pre-fricative tokens (6.1 (b)), male vs. female speakers (6.1 (c)) and early vs. late speakers (6.1 (d)).

- (3) The vast majority of tokens which are located near the TRAP average are DANCE tokens (fig. 6.1 (b)), which shows that the BATH subset transferred to START earlier than DANCE. This is of course in keeping with what we would expect on the grounds of the situation in Australia (cf. section 6.1 above).
- (4) Finally, we note that if the data is broken up into male/female speakers (fig. 6.1 (c)), there seems to exist a preference on the side of the male speakers to retain the more conservative variant, i.e. the tokens that are located in the vicinity of baseline TRAP are predominantly from male speakers.

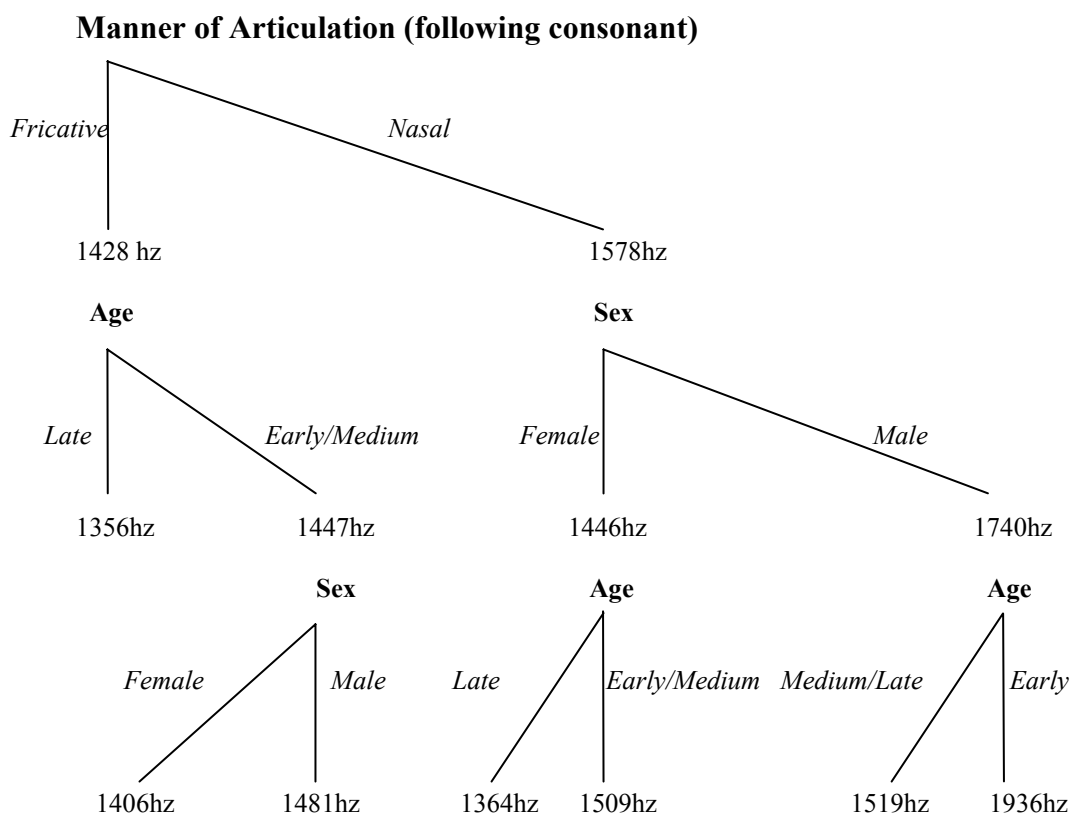


Figure 6.2 – Classification and regression tree analysis of pooled F2 data on the independent variables *Manner of articulation of the following sound/Speaker Sex/Age*. Note that in this figure the length of a given branch is not related to the significance level of the branching.

Therefore, there are clear patterns in the overall distribution of DANCE/BATH with respect to speaker age, speaker sex and lexical subset. Figure 6.2 models these variables in terms of a CART regression tree with the normalised second formant frequency as the dependent variable.

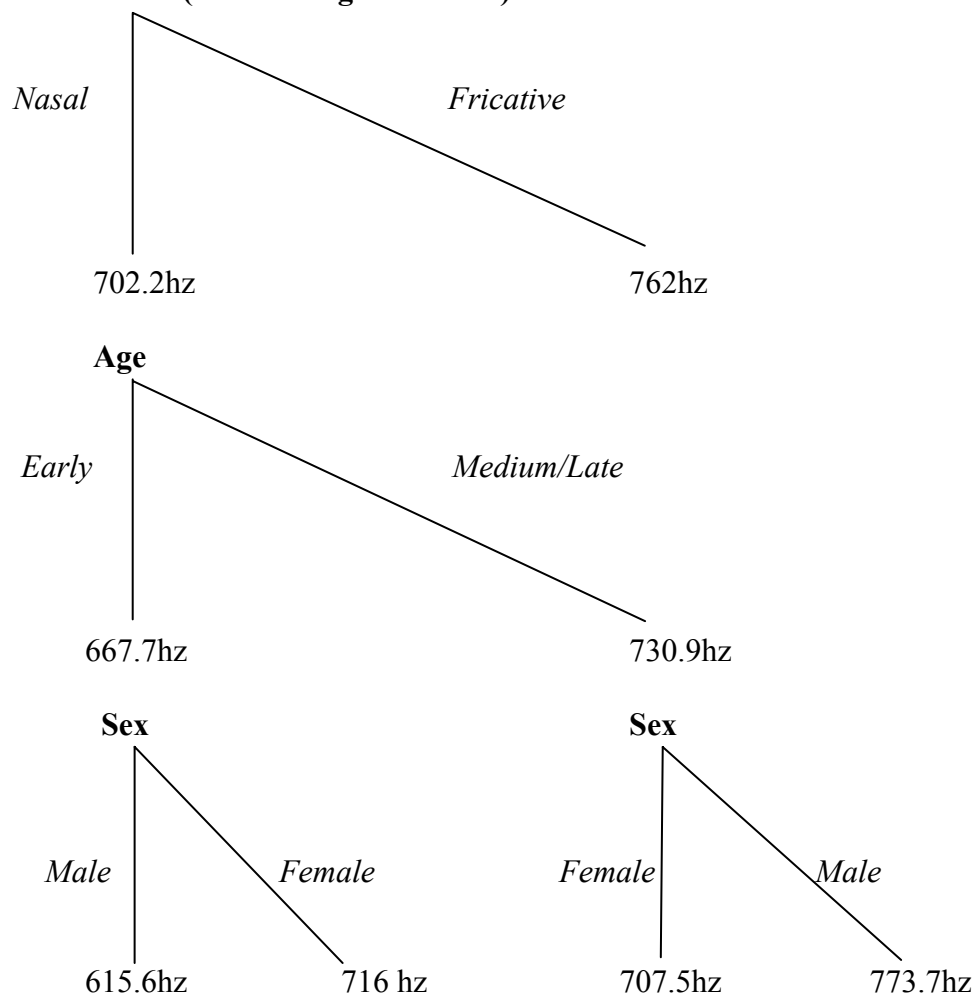
**Manner of articulation (of following consonant)**

Figure 6.3 – Classification and regression tree analysis of pooled F1 data on the independent variables *Manner of articulation of the following sound/Speaker Sex/Age*. Note that in this figure the length of a given branch is not related to the significance level of the branching.

On the whole, fig. 6.2 shows the expected relationships: The main subdivision is in terms of lexical subset, where pre-nasal tokens (i.e. DANCE) show a significantly higher F2 average than pre-fricative tokens (i.e. BATH). Within the BATH branch, there are minor subdivisions along the variables age (where late speakers show a lower F2 average than the other two groups) as well as speaker sex, where males show a higher F2 average than females within the early/medium age branch. Within the DANCE branch, the pattern is reversed in that speaker sex takes precedence over age in predicting the F2 average. Male speakers show a markedly higher F2 average than females (by 264 hz), which seems to be due mainly to the

early male group where the majority of DANCE tokens has not yet transferred to the lexical set of START.

A similar, if somewhat less stratified picture arises if the first formant frequency is regressed on the same independent variables (fig. 6.3). As with the F2 data above, the main subdivision is into lexical subsets, where the average F1 is significantly lower for DANCE words than for BATH words. In addition, the DANCE data is subdivided into age (with early speakers showing lower F1s) and speaker sex. As for the latter variable, however, the pattern is ambiguous: Whereas in the early group it is the males who show lower F1 averages, this pattern is reversed in the other branch (i.e. the late and medium speakers). The subdivision within the medium/late group is rather unexpected, since on the whole, females seem to show more innovative realisations (i.e. higher F1 averages). It should be noted, however, that all tokens within the superordinate branch are clearly START vowels. Therefore, the lower F1 averages in the female terminal branch probably reflect an independent alternation with regard to the openness of the START vowel on the whole whereby medium and late females have a closer vowel than their male counterparts<sup>39</sup>.

## 6.5 Merger by transfer?

An analysis of formant frequency averages alone does not conclusively show whether what we are dealing with is a transfer of tokens from phoneme A to phoneme B or a gradual movement of a number of words (or phonemic environments) from A to B. In this section, we will show that the former is the case. Under the ‘merger by transfer’ hypothesis, we would expect a clustering of tokens around the baseline averages of either TRAP and START (for speakers who have remnants of the earlier pattern) or the target phoneme alone (for speakers who have the modern pattern), but not too many tokens in-between. In other words, we would expect for the most conservative speakers a bimodal overall distribution where each mode

---

<sup>39</sup> Which may still be somewhat unexpected: NZE has been shown to have undergone START fronting (Gordon et al. 2004: p 132), and female speakers were shown to have led that type of sound change. If we associate fronting with opening, we would expect an inverse pattern. If we assume fronting+closing, the present findings are expected. The above-quoted analysis in Gordon et al. assumes height-neutrality under fronting and therefore does not address this question. The present data does not allow for a comprehensive analysis of this phenomenon. At any rate, the effect should probably not be overstated as it applies only to DANCE tokens in figure 6.3 above, i.e. only a rather restricted subset of START.

corresponds to the major predictive factor of the overall distribution (i.e. lexical set, as analysed in section 6.1 above) and a normal distribution around the mean of the target phoneme for speakers with the modern pattern. This seems to be the case.

Figure 6.4 plots the formant frequencies of all DANCE/BATH tokens in the speech of the early male (6.4(a)) and the late female group (6.4 (b)), figure 6.5 shows histogram representations of the same data for each subsample.

Fig. 6.4 (a) shows that the types of the overall distributions in the two groups are different: Whereas the distribution of all DANCE/BATH tokens is approximately normal in the late female group (6.5 (f)), it seems as if there are two modes in the early male group, one at formant frequencies between 1400 and 1500 hz (which corresponds to the late female average) and a minor one at 1800-1900 hz. In addition, the pattern of the latter group is resolvable in terms of lexical set identity, which is suggestive of a transfer between the two phonemes. If this is true, we would expect two potential scenarios in an in-between stage: Either no in-between stage at all, and all DANCE tokens move over to START/BATH at once, or an in-between stage where the DANCE tokens assume a bimodal distribution with themselves (since under this scenario, some tokens would have moved over to START, whereas others remain in place). Under a gradualist (i.e. merger by approximation) scenario we would expect the high F2 (TRAP) mode to maintain its original size, but closer to the BATH mode than was the case at the earliest stage. Figure 6.6 and 6.7 below show histogram representations of DANCE (fig. 6.6) and BATH (fig. 6.7) tokens for all subsamples. Whereas the BATH data shows a similar pattern in the medium female sample as it did in the late one (albeit with somewhat more rigid cut-offs near the maximum), the DANCE data is suggestive: Although token numbers are somewhat small, it does indeed look as if the DANCE set had assumed a bimodal distribution with a prominent node near the overall BATH mode and another one near the original DANCE mode. It should, however, be pointed out that the secondary mode is mainly due to three productions by one individual speaker.

However, there is an interesting complication that shows up especially in the BATH data. In the early and the medium male sample, the column containing the largest number of tokens is in front of the START average, whereas it is backer in the late male sample. While the START mode fronts between the early and the medium stage, this incremental BATH backing does not seem to be an artefact of that, since the medium males show fronted START,

but not fully backed BATH. The female data seems equally confusing: Again, we have fronting of START between the early and the medium stage, and the same change-over in frequency-of-distribution maxima in BATH between the early and the medium female sample. Although the most prominent BATH mode coincides with the START average in the late female sample, the bulk of BATH tokens is located to the left of that mode. It is unclear why BATH should behave in such a way. Coarticulatory effects can probably be ruled out since BATH behaves differentially with regard to the START average across samples. Rather, it seems to be as if BATH behaved independently of the rest of START even for speakers that supposedly show completion of transfer of BATH from TRAP to START. This might imply a gradual sub-process of BATH-backing independently of the transfer from TRAP to START. We will come back to this point in the following section.

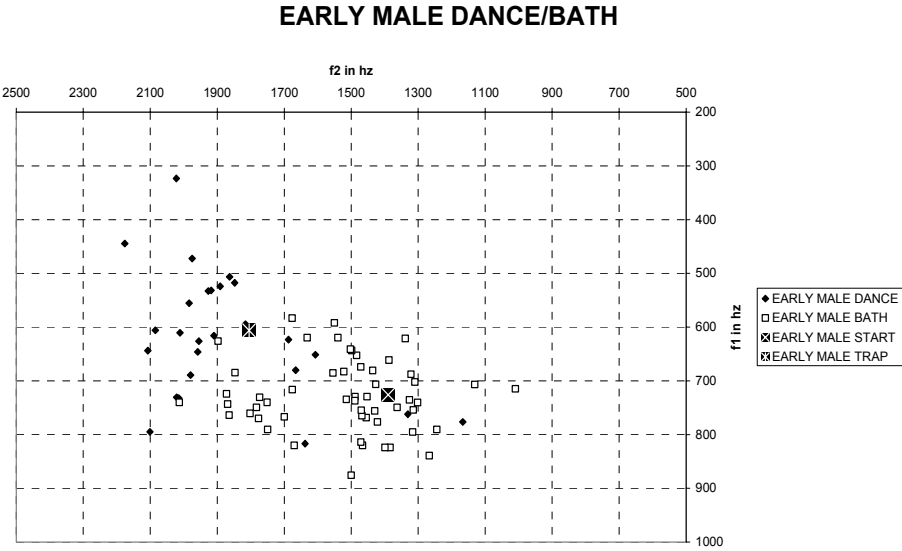


fig. 6.4 (a)

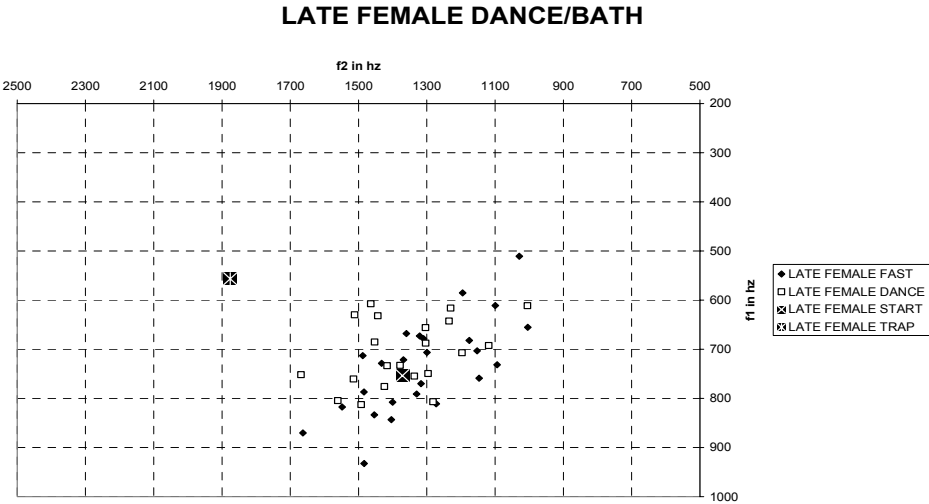


fig. 6.4 (b)

Figure 6.4 – Formant frequency plots of DANCE/BATH against TRAP/START averages of the early male (6.4 (a)) and late female (6.4 (b)) sample.



Early Males - F2 Distribution of potential broad A tokens

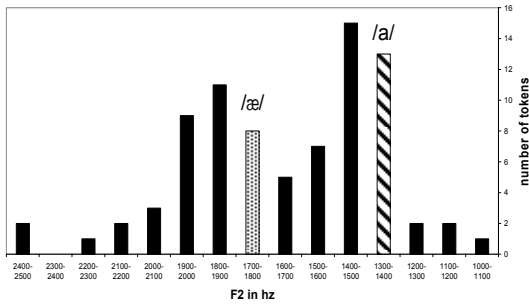


fig. 6.5 (a)

Early Females - F2 Distribution of potential broad A tokens

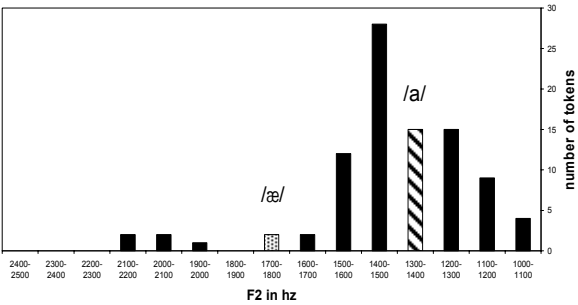


fig. 6.5 (b)

Medium Males - F2 Distribution of potential broad A tokens

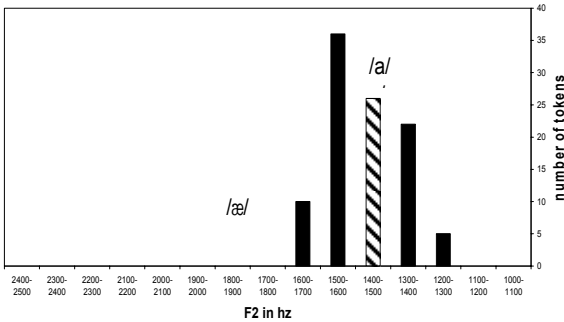


fig. 6.5 (c)

Medium Females - F2 Distribution of potential broad A tokens

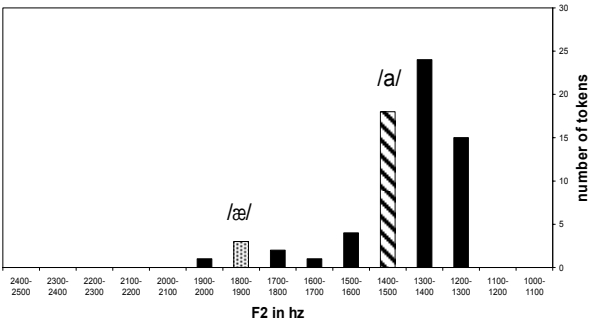


fig. 6.5 (d)

Late Males - F2 Distribution of potential broad A tokens

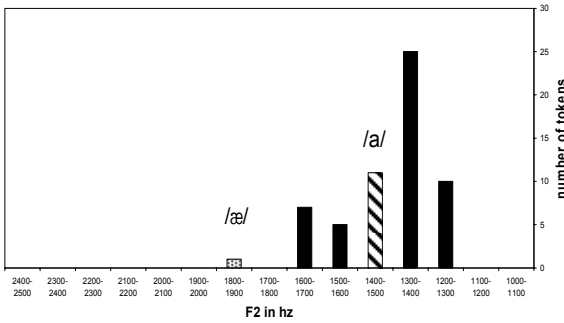


fig. 6.5 (e)

Late Females - F2 Distribution of potential broad A tokens

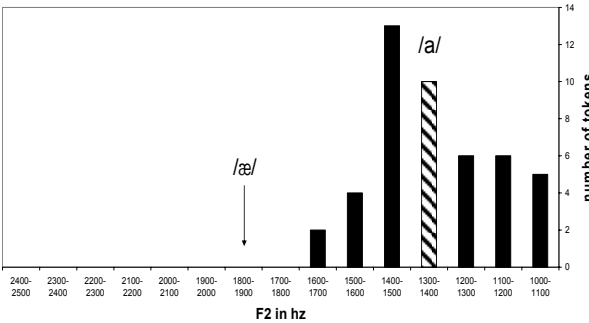


fig. 6.5 (f)

Figure 6.5 – Histogram representations potential of broad A words. Token numbers (y-axis) are plotted against F2 divided up into 100hz steps (x-axis). Also shown are the averages of TRAP and START for each subsamples.

Early Males - F2 distribution of BATH

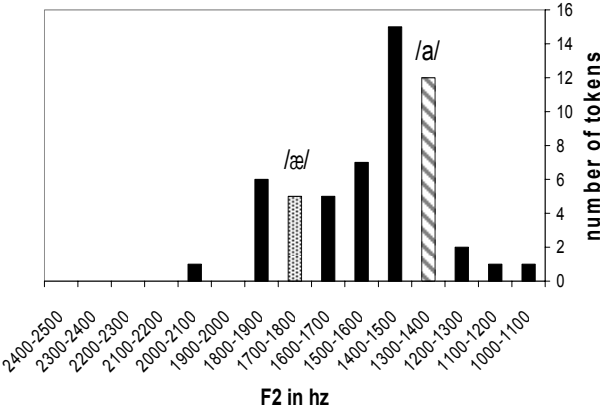


fig. 6.6 (a)

Early Female - F2 distribution of BATH

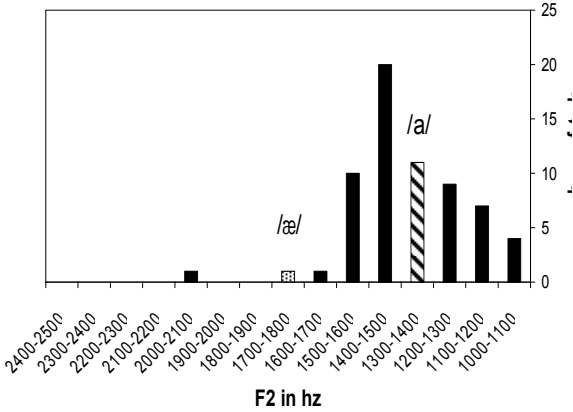


fig. 6.6 (b)

Medium Males - F2 distribution of BATH

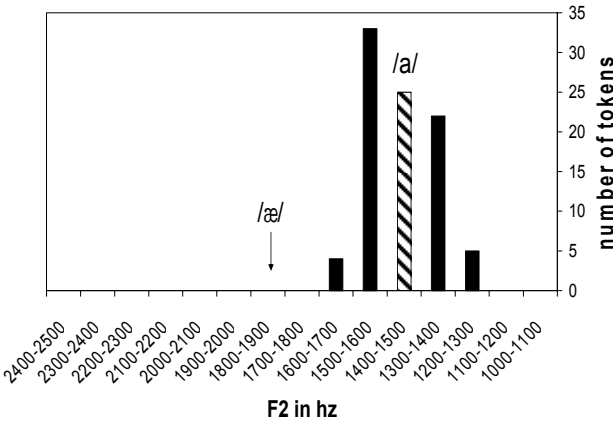


fig. 6.6 (c)

Medium Females - F2 distribution of BATH

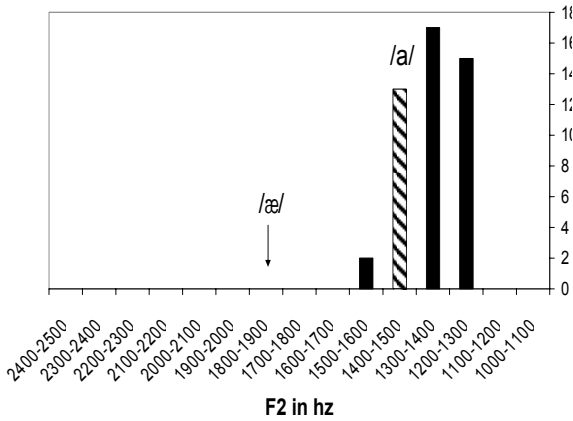


fig. 6.6 (d)

Late Males - F2 distribution of BATH

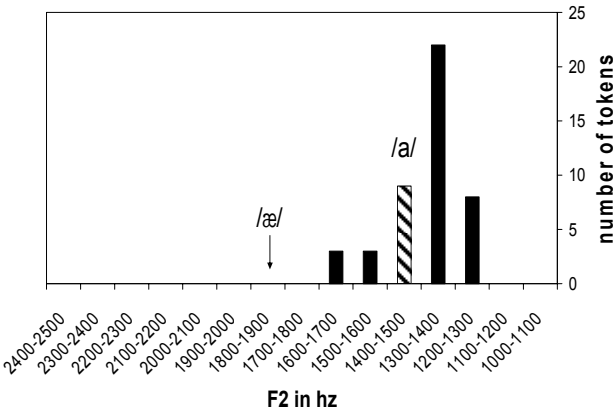


fig. 6.6 (e)

Late Females - F2 distribution of BATH

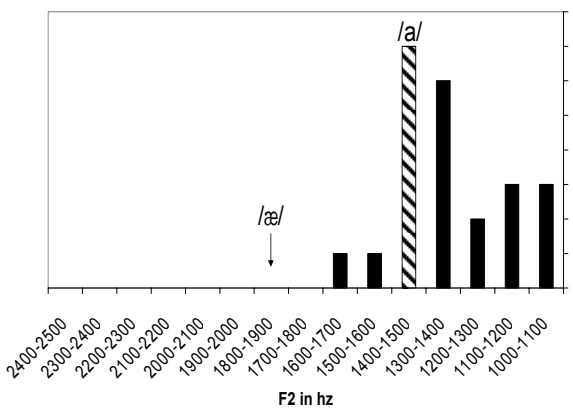


fig. 6.6 (f)

Figure 6.6 - Histogram representations of BATH. Token numbers (y-axis) are plotted against F2 divided up into 100hz steps (x-axis). Also shown are the averages of TRAP and START for each subsamples.

Early Males - F2 distribution of DANCE

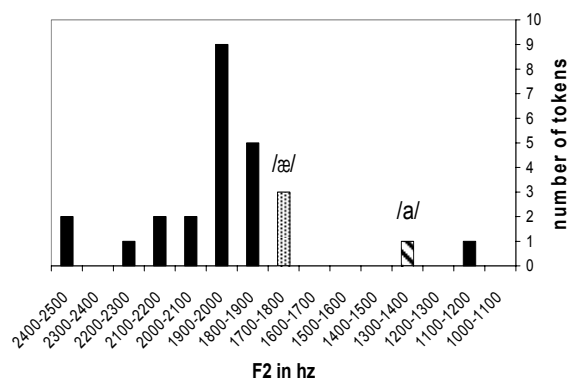


fig. 6.7 (a)

Early Females - F2 distribution of DANCE

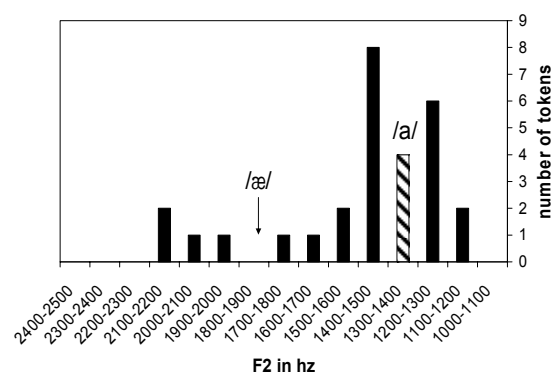


fig. 6.7 (b)

Medium Males - F2 distribution of DANCE

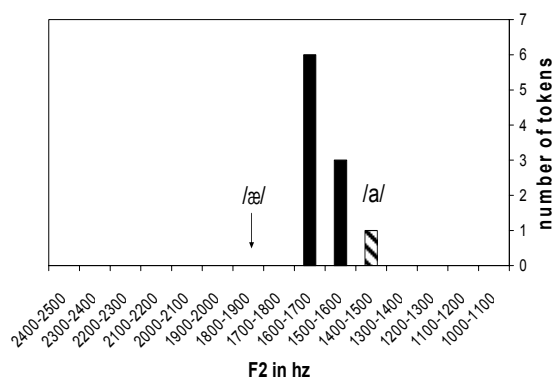


fig. 6.7 (c)

Medium Females - F2 distribution of DANCE

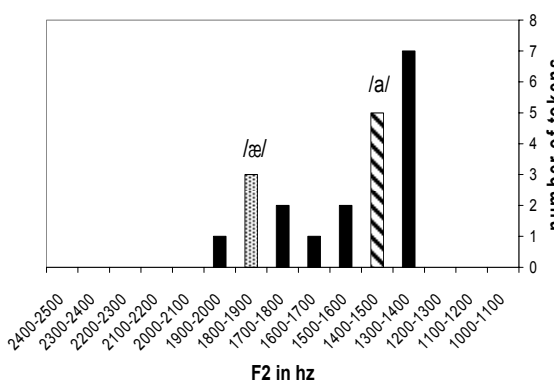


fig. 6.7 (d)

Late Males - F2 Distribution of DANCE

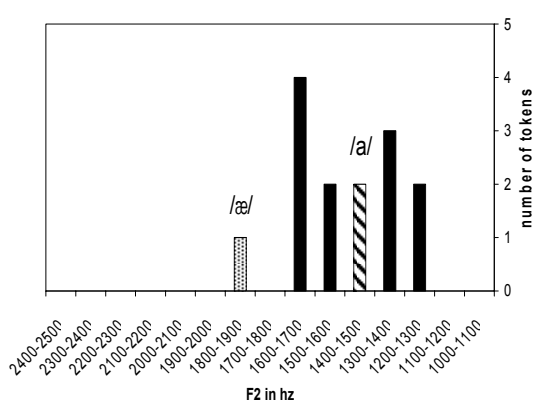


fig. 6.7 (e)

Late Females - F2 distribution of DANCE

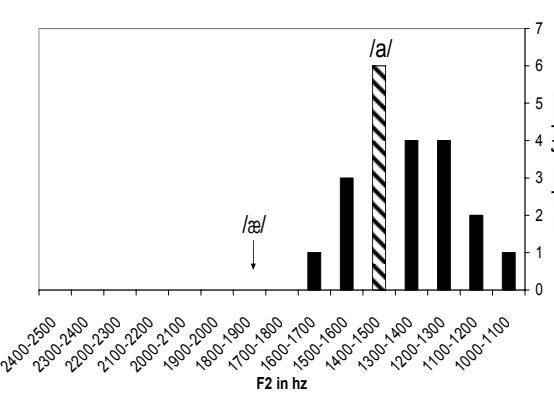


fig. 6.7 (f)

Figure 6.7 - Histogram representations of DANCE. Token numbers (y-axis) are plotted against F2 divided up into 100hz steps (x-axis). Also shown are the averages of TRAP and START for each subsamples.

## 6.6 Vowel duration

As was discussed in the introductory section, the recent history of ME short /a/ involved not only a number of quality changes, but also at least two lengthening processes, the main one being the (originally allophonic) lengthening of TRAP tokens which now comprise the START class in RP and NZE as well as ‘secondary’ lengthening within START as a result of the loss of non-prevocalic /ɪ/ (although, as was mentioned in the introductory section, it is not clear whether that process left any trace in the realisation of vowel length within START vs. DANCE/BATH). These were argued to have come about independently of each other, and we would therefore expect to be able to trace back lengthening and backing separately in an intermediate stage. Table 6.2 below states mean durations of the following classes of vowels: Baseline TRAP before voiced plosives (TRAP1) and voiceless plosives (TRAP2), BATH (no length measurements of START have been obtained in the context of the present analysis), DANCE with a front nucleus (i.e. TRAP, here DANCE1) and DANCE with a non-front nucleus (i.e. START<sup>40</sup>, here DANCE2). In addition, table 6.2 states the significance levels of the differences between the sub-classes as obtained by a Wilcox test.

---

<sup>40</sup> Note that this binary classification has been arrived at on the basis of a concomitant (forced choice) auditory analysis. This analysis, however, concords rather well with the findings from the instrumental analysis discussed above.

		TRAP1	TRAP2	BATH	DANCE1	DANCE2
<b>Mean Duration</b>		106ms	94.8ms	141.4ms	114.5ms	131.6ms
<i>Difference to</i>	<i>TRAP1</i> / _ [+voiced] [-cont]	n/a	<b>p&lt;.01</b>  <b>W = 1515.5</b>	<b>p&lt;.0001</b>  <b>W = 3939</b>	p>.3  W = 703.5	<b>p&lt;.001</b>  <b>W = 1241.5</b>
	<i>TRAP2</i> / _ [-voiced] [-cont]			<b>p&lt;.0001</b>  <b>W = 1994.5</b>	<b>p&lt;.0001</b>  <b>W = 376.5</b>	<b>p&lt;.0001</b>  <b>W = 667.5</b>
	<i>BATH</i>				<b>p&lt;.0001</b> <b>W = 8572</b>	<b>p&lt;.05</b> <b>W = 16139.5</b>
	<i>DANCE1</i> [æ]					<b>p&lt;.01</b>  <b>W = 1023</b>

Table 6.2 – Duration differences and significance levels of pooled data from the Intermediate sample. Bold numbers indicate differences below the .05 level of significance.

The data suggests a rather fine-grained gradation in duration differences. The general direction is clear: The more START-like the sub-set in question is, the longer its duration. That is, the largest duration differential holds between baseline TRAP2 and BATH, while the pre-nasal tokens hold an intermediate position. DANCE1 (i.e. DANCE tokens coded as TRAP) is shorter than DANCE2 (i.e. DANCE tokens coded as START), which in turn are shorter than BATH. The lack of a significant length difference between DANCE1 and TRAP1 would be an intriguing sign that the lengthening of the transferring pre-nasal tokens and their change in articulatory space are not independent processes (at least not for these speakers), but rather inexorably linked. That is, a DANCE vowel which has not yet gone over to the lexical set of START retains the duration associated with an associated non-transferring TRAP vowel (e.g. pre-nasals such as ‘man’). However, there is one complication: According to the analysis outlined in Lass (1990) and Lass (1997), English underwent a further process of lengthening (which Lass terms ‘lengthening II’) of the TRAP vowel (in fact, all vowels) in environments before nasals and voiced stops. This might imply that the lack of a significant length difference between DANCE1 and comparable TRAP items may not necessarily be due to the process as outlined above, but rather to two historically distinct processes: “Lengthening I”

(START, BATH, DANCE) and “lengthening II” (TRAP before voiced stops and nasals). Unless the latter ‘adds up’ to the former, the two would be indistinguishable on a post hoc basis with the data at hand, since no duration measurements of TRAP vowels before voiced fricatives were taken (note that this would be the only viable control variable, as it should be unaffected by ‘lengthening2’. But even so, all DANCE vowels would have been subjected to ‘lengthening II’ at any rate and would be expected to be longer than the control tokens).

One point that can be made to the effect of the independence of lengthening and transfer with regard to the present data is the significantly greater duration of BATH vis-à-vis DANCE2. Since we have seen above that BATH words transferred to the lexical set of START earlier than DANCE words, there seems to have come about a moderate lengthening process within these words, possibly to ‘synchronise’ with ‘proper’ START words (which would have undergone additional lengthening in compensation for the loss of the pre-consonantal rhotic). And since DANCE2 is shorter than BATH, this cannot be explained by phonemic environment (since ‘lengthening 2’ would favour longer DANCE2). The scenario that has been argued for so far, namely one of straightforward transfer of words from one lexical set to another, therefore seems to be somewhat weakened.

In addition to the observation made in section 5.5 above regarding what looks like gradual backing of BATH after the transfer of BATH tokens to START was completed, we have further evidence for gradual sub-processes in the overall transfer.

## 6.7 Lexical diffusion and the role of proper nouns

From the second stage of the overall shift (i.e. the lengthening I phase) onwards, the process became increasingly less predictable. That is, in all accents that underwent the ‘broad A’ shift, the ‘classes’ of words which split off and went over to the START set have been notoriously indefinable on straightforward phonemic grounds. Although some general tendencies can be seen (TRAP lengthened/backed before /r/, voiceless fricatives + another consonant, nasals + consonant with the exception of the velar, and so on), each of these has various counterexamples. Therefore, this process has been accepted as one of the few clear examples of lexical diffusion (cf. the discussion in Labov 1994).

In order to show this with the present data we would need to show that within a subclass that is variable within the corpus or a sub-part of it (i.e. DANCE in the early male sample), the classification of a word as TRAP or START (or its respective position on the F1/F2 plane) should be rather uniform for each speaker. In addition, we would expect lexical identity to be a better predictor of lexical class membership than phonemic environment. Unfortunately, token numbers are too small in the present sample to allow for a statistically valid analysis of these relationships. In addition, the sample of DANCE tokens is considerably paucal: In the early group, 24 out of 53 DANCE tokens are ‘dance’ or some derivative of it. Some points, however, deserve mentioning in this respect. Figure 6.8 below plots all DANCE tokens from both early speaker groups, broken up into two subsets: The lexical item ‘dance’ vis-à-vis all other tokens of DANCE.

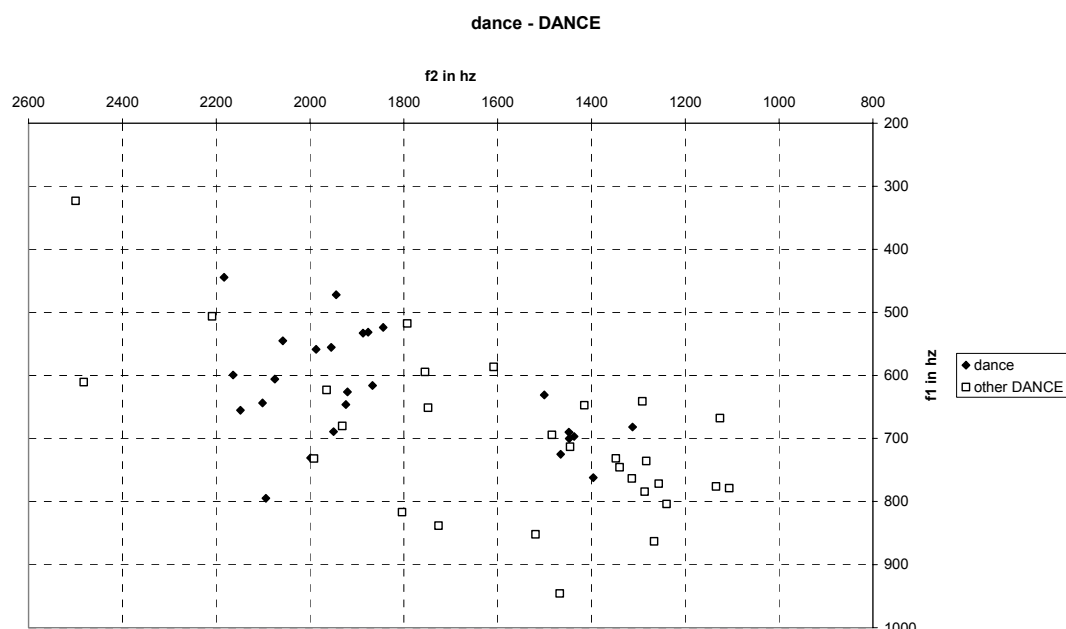


Figure 6.8 – Formant frequency position of the lexical item ‘dance’ vs. all other pre-nasal tokens which are subject to the broad-A rule (only early speakers are shown).

It is clear that on the whole, ‘dance’ is rather more conservative than most of the overall lexical set (a Hotelling-Lawley trace score of 0.35,  $F \sim 8.85$ ,  $p < .001$ , indicating a significant separation between the two distributions). The majority of DANCE tokens near the TRAP

average are instances of the word ‘dance’, of which only 7 have moved over to START for these speakers. In addition, 6 out the 7 innovative ‘dance’ tokens come from one and the same speaker (Mrs A. – She does not have any ‘dance’ tokens realised with the TRAP vowel). The TRAP tokens of ‘dance’ come from 3 other early speakers (2 males and 1 female).

There also seems to be some sporadic evidence that proper nouns retain the TRAP variant more readily than other words. Figure 6.9 below plots the formant frequencies of three speakers from different stages in the intermediate period. 6.9(a) shows Mrs. A’s (early female) DANCE/BATH tokens compared to two instances of the utterance ‘Castle Street’, 6.9 (b) shows Mrs S’s (medium female) DANCE tokens divided up into proper nouns (either ‘Alexander’ or ‘Alexandra’) compared to her other DANCE vowels (all instances of the word ‘dance’), 6.9(c) shows a token of the proper noun ‘Alexandra’ in the speech of Mr BA (late male) compared to his remaining DANCE vowels.

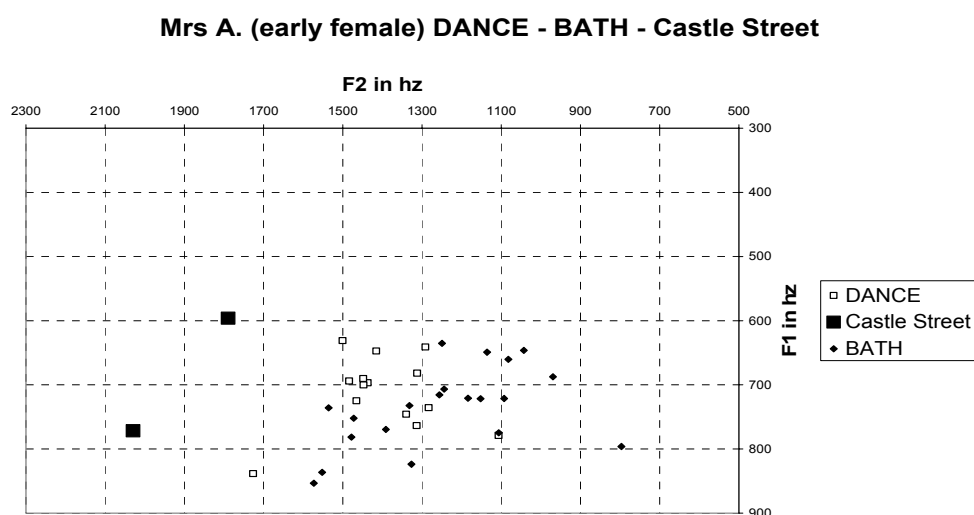


fig. 6.9 (a)





fig. 6.9 (b)

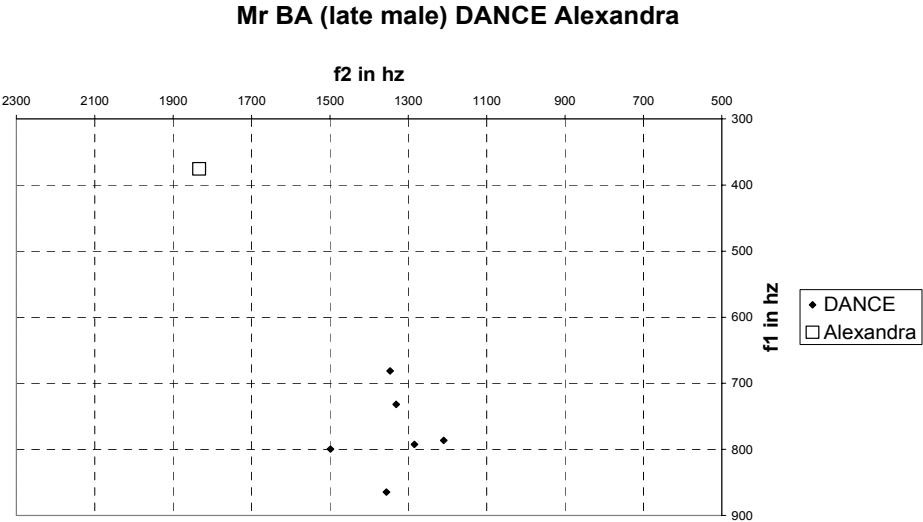


fig. 6.9 (c)

Figure 6.9 – Formant frequency plots showing conservative (i.e. TRAP) realisations of place names in the speech of an early female (6.9 (a)), a medium female (6.9 (b)) and a late male (6.9 (c)) speaker, respectively.

However, as the above data nearly exhausts the available data on proper nouns, it seems prudent to remain cautious regarding the inhibitive role of proper nouns in the DANCE/BATH transfer (although the same pattern has been found in the Mobile Unit data. M. MacLagan, personal communication).

## 6.8 Conclusion

From the analysis presented in this chapter we can conclude that

- During the Intermediate Period the transfer of DANCE/BATH tokens from the lexical set of TRAP to START was completed. Whereas the early speaker groups still show the front vowel in a number of tokens, the late speakers have exclusively START.
- The pre-fricative vowels (BATH) moved over to START earlier than the pre-nasal vowels did. Even in the early samples, the front vowel survives only in a small number of words if followed by a fricative. Pre-nasal tokens, on the other hand, were still commonly produced with the TRAP vowel in the speech of the earlier speakers, especially the males.
- The distributional properties of the DANCE vowel strongly suggests that the process (at least at this late stage) is a merger of transfer rather than a merger by approximation.
- The transfer of tokens may well be subject to lexical diffusion, as the distribution of one lexical item (here ‘dance’) does not necessarily conform to the overall distribution (i.e. DANCE). However, token numbers are clearly too small to allow for a satisfactory analysis of this phenomenon with the present data.
- There is some evidence that proper nouns tend to behave rather more conservatively than lexical nouns since for some speakers they retain the TRAP vowel whereas comparable lexical nouns have the START vowel.
- Vowel length is largely a function of lexical set membership, and the historical independence of lengthening and backing is not borne out in the present data.
- However, in addition to this, there are differences in phonetic detail between some of the subsets that suggest a small amount of gradual sound change in addition to the transfer process.

On the whole, the above conclusions are in keeping with Gordon et al.'s analysis of these variables in the Mobile Unit where TRAP realisations of DANCE/BATH were slightly more common than is the case for the early intermediate speakers. However, as far as the role of speaker sex is concerned, the findings are different: Whereas in Gordon et al.'s analysis male speakers born after 1875 lead the change toward broad-A, it is the female speakers who are ahead in the present sample. I have no explanation for this discrepancy other than to point to the fact that during the period in question there seem to have been conflicting views regarding the 'appropriateness' of broad-A in DANCE/BATH words (for details cf. Gordon et al. 2004: p127 - 29): Whereas the broad-A realisation seems to have been (at that stage at least) adopted as the prestige variant by some, others were still aware that the front vowel is etymologically 'correct'. Hence, this is one of the few examples of an alternation where self-proclaimed language watchdogs permitted either variant (cf. A. Wall's comment quoted in Gordon et al. 2004: p. 127). Since males are commonly ahead in stigmatised sound change whereas females lead changes toward prestige variants (to considerably simplify the sociolinguistic generalities), we might here be confronted with a shift in prestige whereby the flat-A variant was the prestigious one in the late Mobile Unit period (hence males were ahead in shifting toward the new variant), but a 'rustic' (non-prestigious) form at the onset of the Intermediate Period.

## 7.

## The Short Front Vowels – An Exception to Directionality in Chain Shifts?

## 7.1 Introduction

This chapter will discuss a number of points of more general import that the data analysed above may help to shed light on. More specifically, we will discuss Labov's theory of vowel shifts and mergers and its fit to the data from the Intermediate period. As it stands, the developments in the short front vowels in the Southern Hemisphere varieties pose a significant challenge to the unified theory of directionality in chain shifts as proposed by Labov (1994/2001). The remainder of this thesis will take two divergent approaches to this problem: In the present chapter we will review Labov's model of chain shifts as well as more general views of how vowel spaces are structured. Both phonetic and phonological arguments will be revisited. Section 7.2 will outline and discuss the notion of 'tense' and 'lax' vowels as well as that of subsystems within a total set of vowels. Section 7.3 will outline the theory of directional vowel change within and across subsystems as advocated by Labov. 7.4 will incorporate the data analysed in the present study into a Labovian view of vowel spaces. More specifically, I will attempt to outline a view of the vowel space in English that maintains the general ideas regarding directionality of vowel changes while allowing for an incorporation of the seemingly 'unlawful' developments in the short front vowels in Southern Hemisphere varieties of English.

The following chapter 8 then takes a more radical approach to the problem by arguing that some of the concepts advocated by Labov, such as 'tracks' and 'subsystems', may be redundant and that evolution in vowel systems may be better described along a single spatial dimension, namely *distance between vowels as a function of a number of different phonetic dimensions*.

## 7.2 Vowel types and subsystems

Neither the classification of vowel systems as wholes, nor that of their members is entirely straightforward, and a number of points have to be discussed in this context. The questions that are relevant in the current context are as follows:

- (1) Are vowel spaces segmented (in general)?
- (2) Is the vowel space of English segmented?
  - (2) (a) If so, how many types of vowels are there?
  - (2) (b) Which vowel belongs to which type?
  - (2) (c) How do we label different types?
  - (2) (d) If there are different types, what is the relation between vowels across and within subsets?

### *Are vowel spaces segmented?*

This relates to whether vowels in general fall into different types, and how we can assess the notion of types. That there are, in certain languages at least, different types of vowels has long been recognised<sup>41</sup>. The criteria for distinguishing between types, however, are less uniform and can be either phonetic, phonotactic, morphophonemic or historical. I discuss each of these in turn.

---

<sup>41</sup> This goes back to at least the orthographisation of Latin, which has the same grapheme for what were clearly different vowel phonemes (such as *honor* with a long vowel in the first syllable, and a short one in the second one). In later scribal practice, these were often distinguished by applying a superscript macron to the long vowel. Note, however, that the actual grapheme was maintained.

*(a) phonetic criteria*

Phonetic distinctions of vowel types are probably the most traditional ones and have aimed at establishing vowel types in term of *quality* and *quantity*. The former relates, for example, to the monophthong/diphthong (or rather, polyphthong) difference or to that between nasal vs. oral vowels, whereas the latter recognises differences in length, i.e. duration, between vowels. For example, in a language that has both monophthongs and diphthongs, the monophthongs would form a coherent type (or subsystem) that can in some sense be meaningfully contrasted with that of the diphthongs on phonetic grounds (i.e. in this case, all members, and only those, of the former set show articulatory movement throughout the duration of the segment). The problem with these is that a purely superficial (in the technical sense) differentiation of vowels into types may actually be misleading in that phonetic form may not necessarily correspond to linguistic structure, i.e. that linguistically meaningful distinctions may cut across these phonetic types. In addition, deciding upon classification of some item may not be straightforward in each particular case. For example, the monophthong/diphthong distinction may be hard to ascertain given that coarticulation with consonants occur even in monophthongs, which might bring about non-steady articulations of these segments. And since this definition is based upon surface criteria only, distinguishing between monophthongs and diphthongs might become a problem in principle<sup>42</sup>.

*(b) phonotactic criteria*

This relates to the question of which phonemic environment a given vowel can occur in, and whether we can make generalisations of occurrence restrictions across the entire set of vowels. This approach toward classifying vowel systems is therefore structural rather than phonetic, and is preferred in more recent approaches (at least since Trager/Smith 1957). An example of a subdivision of vowel systems that is phonotactically motivated would be that of Swedish vowels into long and short ones. Although the labelling is ultimately based on

---

<sup>42</sup> Although vowels in isolation can be said to constitute test cases. Whereas this works well for English (since English allows for most of its long vowels/monophthongs to occur as lexical items (cf. 'I', 'are'), or at least pseudo-lexical items ('ooh', 'oi'), this might be a problem in languages that do not allow for syllables which only consist of a nucleus. There, vowels in isolation would be linguistic non-entities and may not necessarily be informative regarding articulations in actual language.

phonetic criteria, this subdivision is strengthened by the fact short vowels, and only short vowels, can occur before consonant clusters or geminates, that is, Swedish syllable structure disallows light as well as superheavy syllables. Examples from English include the restriction of schwa to unstressed syllables as well as the famed tense/lax distinction, which will be discussed in more detail below.

*(c) morphophonemic criteria*

Ever since Chomsky & Halle (1968), morphophonemically regular alternations between surface forms have become popular test cases for elucidating relations between vowels. The gist of this type of analysis is that in certain types of morphological operations, vowels alternate with each other in a regular way, and therefore form linguistically meaningful pairs (examples of this type of process will be discussed below).

*(d) historical criteria*

A further indicator that there exist different types of vowels comes from historical evidence. More specifically, it has been shown that certain patterns of interaction between vowels in chain-shift relationships are inexplicable if vowel spaces are treated as continuous spatial arrangements of vowels without further substratification. To give an example: A number of Germanic languages (English, German, Dutch, to some extent also Swedish and Norwegian) underwent similar processes of presumably interrelated vowel movements. This can be diagrammed as follows (I restrict myself to the non-low front vowels here, since the functional relationships seem clearest in these vowels):

/e/ > /i/ > /ei/ ~ /ai/

In brief, a mid-vowel raises by one slot, and a high vowel diphthongises. Whereas German and English have undergone this process, languages like Low German or Danish have not. Swedish and Norwegian show this development in the back vowel series only. Across the Germanic languages, this yields regular correspondences between /e/ (in languages that have

not undergone the shift) and /i/ (in languages which have), and /i/ and /ai/, cf. English ‘my, mine’ and German ‘mein’ as opposed to Swedish ‘min’, Low German ‘min’, or English ‘see’ vis-à-vis Swedish ‘se’ [se:]).

The assumption that each step of the above chain of processes is functionally related to another other one rests on the finding that such movements (which, in the Germanic case at least, are rather more comprehensive than the above schematisation indicates) come about in relatively short periods of time. In addition, in a number of cases it can be shown that if one of the subprocesses does not occur in a given language or dialect, the following ones do not occur either. Lass (1992) makes an elegant argument in this context by showing that in a number of traditional dialects of Northern England, ME /i:/ shows a diphthongised nucleus whereas ME /u:/ remains monophthongal. He furthermore points out that this is the case only in those dialects that previously underwent fronting of ME /o:/ to /ø:/, later /e:/. Since there is no back mid-vowel in these dialects to ‘push out’ ME /u:/, it does not move out of the way).

This type of evidence has therefore been assumed to be indicative of ‘chain shift’ types of processes, where movements that occur in one vowel affect those of another vowel. The problem that arises is how to express constraints onto the interaction patterns between elements of a shift in spatial terms. There is something uncannily local about any two successive steps in most chain shifts, in that the movement of some vowel phoneme /X/ causes reactionary movement in another phoneme /Y/ if both phonemes are located in similar regions of the vowel space. Therefore, shifts of this type have been claimed to involve interaction patterns between adjacent vowels. However, this is strictly speaking not true if adjacency is purely a function of location in articulatory or acoustic vowel space. This is because in Germanic languages (in their ‘Middle’ stages at least) there is one vowel ‘between’ /e/ and /i/ (i.e. ‘short i’, /ɪ/, KIT, or whichever designation one might prefer) which on the face of it ruins the generalisation of vowel shifts operating in a strictly local, stepwise fashion. On the basis of this insight, it has been claimed that (in the Germanic case) the chain shifts diagrammed above affect only the ‘long vowels’, i.e. historical change is sensitive to some kind of systemic subdivision within a given vowel system. This insight was elaborated in Labov (1994), and will be discussed below.



*Is the vowel space of English segmented?*

Bearing in mind the above arguments for the existence of segmentation in vowel systems in general, we might ask whether this applies to the vowel space of English. Going through each of the above criteria it seems clear that:

- (1) *phonetically*, there are both monophthongs and diphthongs, round vowels and non-round vowels, as well as long vowels and short vowels. However, it is less clear whether such distinctions can be taken as the basis of a principled subdivision into meaningfully different sets. For example, although in a dialect such as RP there exist clear instances of monophthongs and diphthongs at each end of the continuum, there are borderline cases where this is less clear (e.g. the vowel in FACE). In addition, such articulation-based dichotomies are rather prone to bleaching if we look across dialects and historical periods (where, for example, the vowel in NEAR is a long monophthong in AusE, a centring diphthong in RP, and a short monophthong in GenAm).
- (2) *phonotactically*, there are vowels that can occur in certain environments, but not in others: It is well known that a number of vowels can not occur in open stressed syllables in English (the vowels in KIT, DRESS, TRAP, STRUT, LOT, and FOOT). In addition, this difference is comparatively stable across different varieties of English (excluding varieties with completely different phonotactics such as Scottish, cf. Aitken (1964)) and may therefore be a much better basis of sub-classifying the vowels of English. Further co-occurrence restrictions comprise, for example, the restriction of schwa to unstressed syllables and the fact that velar nasals can only occur after the ‘short vowels’, i.e. those given in parentheses above.
- (3) *morphophonemically*, there are regular alterations between subsets: In trisyllabic shortening, for example, the pervasive pattern is for vowels such as the ones in MOUTH, PRICE, FLEECE, GOOSE, and others to be changed over to those in STRUT, KIT, DRESS (among others, cf. the discussion in Chomsky and Halle (1968)) in certain derivational paradigms (cf. divine – divinity, pronounce – pronunciation).

- (4) *historically*, vowel movements of more than one vowel suggests the existence of subsystems: The Great Vowel Shift, for example, raised long monophthongs by one slot and diphthongised long high monophthongs while leaving the short monophthongs entirely unaffected.

We therefore have reasons to believe that there do in fact exist different types of vowels in English. It would be convenient if the above criteria yielded consistent results in an analysis of English vowels. This, however, is not the case. Rather, classifications arrived upon one of the above criteria may in certain cases run counter to that based on some other criterion. For example, the monophthong-diphthong division clearly cuts across the phonotactic division in that the phonotactic classification identifies two classes of monophthongs (i.e. ones that can occur in open stressed syllables, and ones that cannot), whereas it does not impose any restrictions onto the distribution of diphthongs. The historical interrelations between the vowels of English seem to match the phonotactic classification better than the phonetic one, since long vowels (i.e. the phonotactically unrestricted ones) alternate (i.e. push out, pull in, merge with) with the diphthongs, and not with the other monophthongs. We would therefore have to ask

*How many (essentially different) types of vowels are there, and which vowel belongs to which type?*

Table 7.1 below lists all vowel phonemes of RP and states their classification in IPA, Wells' (1982) lexical set names, and Trager/Smith (1957) transcription. These different transcription techniques already make statements regarding both the nature of the segmentation as well as the designation of each of the subsets, as will be discussed in more detail below.

Lexical set	IPA	Trager-Smith <sup>43</sup>
FLEECE	/i/	iy
KIT	/ɪ/	y
DRESS	/ɛ/	e
TRAP	/æ/	æ
STRUT	/ʌ/	ə
START	/ɑ/	ah
LOT	/ɒ/	a
THOUGHT	/ɔ/	ɔh
FOOT	/ʊ/	u
GOOSE	/u/	uw
NURSE	/ɜ/	n/a
PRICE	/aɪ/	ay
MOUTH	/aʊ/	aew
CHOICE	/oɪ/	ɔy
FACE	/eɪ/	ey
GOAT	/oʊ/	əw
NEAR	/ɪə/	n/a
SQUARE	/ɛə/	n/a
CURE	/uə/	n/a

Table 7.1 – Transcription of Standard English vowels in IPA, Wells (1982) and Trager/Smith (1957).

Whereas the lexical set name approach does not divide up the vowel space into subclasses, the other two approaches do: The IPA transcription essentially posits two ‘kinds of vowels’ for English, (1) vowels transcribed with a single symbol and (2) vowels transcribed with two symbols. This corresponds to the conventional monophthong-diphthong distinction and therefore captures an articulatory fact about English rather than a systemic one<sup>44</sup>. This

<sup>43</sup> Trager & Smith describe a vowel system of the GenAm type, where there are no phonemic centring diphthongs.

<sup>44</sup> Note that this is not the only kind of transcribing the vowels of English in IPA script. It is at least as common to indicate length on long monophthongs by adding the symbol /:/ to a vowel, which, if the symbols as given in table 1 above are used, splits up the vowel inventory of English into three types rather than two.

exemplifies the main problem of this type of transcription: Do we want to capture facts about articulation of vowel phonemes in a *phonemic* transcription at all? If so, how much?

The Trager-Smith approach to some extent does away with this problem by focussing on the structural properties of the vowels. Here, vowels are either simple or complex segments (the argument for why complex segments are indeed complex is rather complex in itself and need not be reiterated here). Again, the familiar patterning arises whereby the vowels in KIT, DRESS, TRAP, STRUT, LOT and FOOT contrast with all other vowels. Therefore there seems to exist in English a two-way distinction in the vowel system that is hard to grasp in articulatory terms. It may therefore be worthwhile to briefly discuss what an appropriate designation of the two types of vowels might be.

*How do we label the two sets of vowels in English?*

Although naming something may be regarded as being of no great importance to the content of the entities thus named, it would still be necessary to establish an agreed-upon nomenclature. In addition, it is a widely observed tendency to assign ‘thinghood’ to entities on the basis of the labels they carry, and mislabelling may therefore have unwanted consequences such as the iteration of wrong or inaccurate definitions and interpretations. As for the case at hand, it would be desirable to label the two sets in a way that (a) distinguishes between them in the first place and (b) captures as much as possible about their ontological content. That this is not straightforward is reflected by the different approaches taken in the literature, which can be summed up as follows (In the following discussion I will refer to the vowels in KIT/DRESS/TRAP/STRUT/LOT/FOOT as ‘set 1’, and to the others as ‘set 2’):

- The traditional view, which labels the two sets ‘long’ and ‘short’ vowels, respectively. The problem with this is that ‘length’ is too intimately associated with actual phonetic duration, which in turn heavily emphasises a dimension which is a) not directly related to phonological structure and b) highly dependent on factors other than the phonological status of a given vowel, such as voicing of the following consonant or vowel height.

- The structural view, which goes back to Trager & Smith's analysis of vowels in American English and recognises two sets of vowels, 'simple' and 'complex'. The simplex elements occur in what I termed 'set 1' vowels above, 'set 2' vowels are formed by adding one of three types of glides to the simplex vowels (the glide elements are /h/, /y/ and /w/).
- The articulatory view, which labels the two sets 'tense' and 'lax'. This is of course the terminology advocated in Chomsky and Halle (1968), whereby the vowels of set 1 are labelled 'lax' and those of set 2 'tense'. The problem with these terms is as follows: First of all, it is not clear whether there is any articulatory correlate to the notions of 'tenseness' and 'laxness', it rather seems that this is a purely structural feature, i.e. a 'contentless dichotomiser' (Lass 1976). Should one bother? Probably yes, since all other features given in SPE do have rather straightforward articulatory correlates (i.e. +/- high/low refer to openness, +/-round to lip rounding, +/-back to tongue position along the (quasi-)horizontal). A feature-based phonology can thereby map onto articulation in a rather straightforward fashion along these dimensions, but not along the tense/lax dimension. Within a framework that derives concrete surface structures from abstract underlying representations, this surely must be a problem. In addition, given the view advocated by Labov (cf. below), it is not clear to what extent 'tense/lax' are actually atoms, or whether they might be decomposable into smaller entities.
- The comprehensive view as advocated by Labov, which also labels the two sets as 'tense' and 'lax' vowels, but explicitly states that these are 'cover terms' rather than articulatory specifications, and that these cover terms are supposed to cover the whole range of differentiating factors without emphasising any one of them in particular. In brief (this point will be discussed in more detail in the following sections), Labov seems to subsume under these labels vowels that behave differentially in various domains, i.e. as historical entities, as phonemes, as physical articulation, et cetera.
- In addition, Labov divides up the vowel space of systems that show the tense/lax distinction as consisting of peripheral and non-peripheral tracks, where tense vowels

are located on peripheral tracks, whereas lax vowels are located on non-peripheral tracks. This essentially grants priority to the quality distinction between the two vowel sets. In addition, if I read Labov correctly, this idea also introduces a differentiation between vowel space and vowels, or between a potential vowel space and the actual set-up of its members. At least this is my interpretation of statements such as

*In chain shifts, tense nuclei rise along peripheral tracks* (Labov 1994:176)

That is, there seem to be two primitives in the description of vowel systems, namely the type of vowel (tense/lax, or long/short) and some spatial structuring which seems to be to some extent have independent reality. Otherwise, a statement such as the above one might simply be rephrased as ‘peripheral vowels rise’ (which is indeed the original statement of the above principle in Labov (1994:116)). However, Labov makes it explicit that peripherality is a function of spatial relationships between members of a given vowel space, which in my opinion nullifies the notion of independent tracks. I will return to this issue below.

Although the details of each of these accounts differ, we can also observe some degree of common ground between them: First of all, there seems to be agreement that there are ‘different kinds of vowels’ in English. In addition, once we abstract from the purely phonetic level, the number of kinds seems to be two: One set consisting of nuclei that are phonetically short, show little or no articulator movement throughout their articulation, cannot occur in open stressed syllables, and alternate with their complementary counterparts in certain derivational paradigms. In addition, they tend to be located in less peripheral regions of the vowel space. These points are all statements about the properties of the two kinds of vowels *as entire sets*, and ignore one important question that has not received as much attention in the literature as it should have (with the exception of Lass (1976: ch. 1), on which the following discussion is based):

*If there are different types, what is the relation between vowels across and within subsets?*

This relates to the question whether the members of each set of vowels have some kind of ‘counterpart’ in the respective other set, or whether they are unordered wholes in this regard.

Why would one assume that there is some kind of pairing across sets? There are a number of possible reasons, including (1) regular phonological alternations between specific cross-set pairs in some language, or English. (2) One set being a structural mirror image of the other one, once we normalise for the factor that distinguishes the two sets (i.e. the structural relations within one set are exactly the same as those in the other one). (3) Articulatory/acoustic similarity of pairs of vowels across sets. (4) Spelling, which uses (in some instances at least) the same symbols for pairs of vowels.

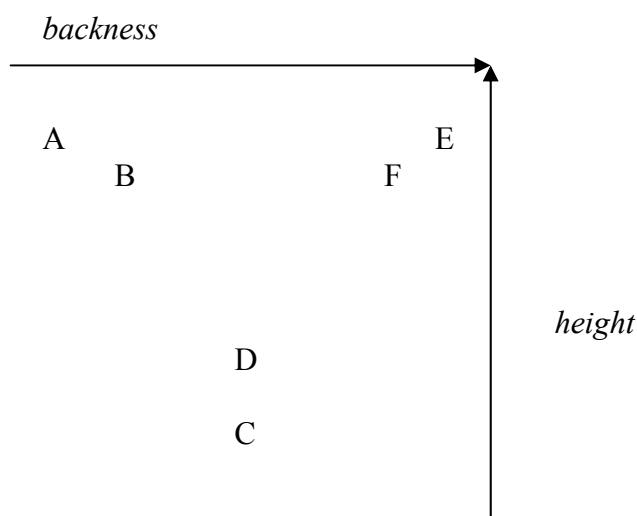
The first point relates to processes of a type that is nicely illustrated by Swedish. In this language, syllable weight is constrained in such a way as to exclude light syllables as well as superheavy ones<sup>45</sup>. Since Swedish allows both long/short vowels as well as single consonants and consonant clusters/geminates, one finds essentially two types of syllables: Ones with a branching nucleus and a non-branching coda (e.g. ‘stol’ /st[[uμ:μ]lμ]/), and ones with a simple nucleus and a branching coda (e.g. ‘bild’ /b[iμ[lμdμ]]/). A case for pairings across vowel systems can be made on the basis of regular alternations in the formation of adverbs. These are formed by adding an alveolar stop /t/ to the adjectival form. As such, the added element would bring about illicit syllable types if the base form is of the type long vowel+simple coda (e.g. ‘stor’ /stu:r/), since one additional mora is added to the coda. What happens in Swedish adverbial derivation is that the long vowel in such cases is supplanted by a short vowel in a rather straightforward way, i.e. /i/ is substituted for /i:/, /ε/ for /e:/, and so on. The question is whether one can make a similar case for English, and what the pairing would look like. On the basis of SPE type of alternations (cf. *divine* – *divinity*) one might posit that this is indeed the case, however, there is something rather inexplicable about these alternations from a synchronic point of view: Why should a given short set 1 monophthong alternate with a set 2 element that is rather different from an articulatory point of view? Of course the answer for this is ultimately based on historical contingencies, since trisyllabic shortening (as a historical process) clearly preceded the GVS (which raised non-high long vowels and diphthongised high long vowels), and that in a pre-GVS system, this type of process makes much more sense (and is rather more similar to the Swedish case above, where the phonological and the vowel-space criteria concur). In addition, since the trisyllabic shortening rule reflects a

---

<sup>45</sup> I adopt the weight assignment advocated in Lass (1984), where light syllables have a branching in neither the nucleus nor the coda on the mora level. Heavy syllables have a branching in either the nucleus or the coda, superheavy ones in both. E.g. in English, ‘bit’ would have a light syllable, ‘beat’ and ‘best’ would have heavy ones, and ‘beast’ a superheavy one.

fossilised pattern of a sound change that was productive at some given point in time, it is of course blind to sound changes that occurred afterwards (which yields non-biunique patterns such as profound – profundity vis-à-vis doubt – dubitable).

The second point relates to the arrangement of vowels in articulatory (or acoustic) space and is fundamental to the Labovian theory of vowel change. That is, do we observe patterns whereby similar relations between vowels of a given subset hold in both sets? Assume an idealised vowel space as diagrammed below, where (A) – (F) represent hypothetical vowels:



How would we subgroup these vowels? It seems clear that in some sense, the vowel space outlined by {BDF} looks rather similar to that outlined by {ACE}, in that the former is a miniature version of the latter. If we observed each of them in isolation, we would probably arrive at similar feature specifications independently (that is, both A and B might be specified as [-back] and [+high], both E and F would be [+back] and [+high], and so on). Furthermore, no other possible subgrouping achieves this type of symmetry<sup>46</sup>. We therefore have reason to state within-system dependencies as properties of both subsystems, and since the number of members is the same in this case, there are no indeterminacies (i.e. B is to D as A is to C, *and*

<sup>46</sup> The question is whether this is pure aesthetics or whether there is something more fundamental about these types of symmetry judgements. One might also say that when looking at Orion (i.e. the stellar constellation) that ‘no other subgrouping achieves the image of a figure with a bow’. I am not sure about this, but will briefly return to this in the final section of this chapter.



A is to B as D is to C. This seems to be the requirement of meaningful pairing across subsystems).

Again, we might wonder whether there are any clear cases of spatial mirroring in real vowel systems of any kind, and in English in particular. The case for subsystems can be made rather straightforwardly in cases like Finnish, where the two sets are identical in spatial terms and distinguished only on the basis of further phonetic properties (such as length in Finnish). Similarly, in languages such as German or Swedish we can probably identify two subsystems since the conditions described above hold rather well in these languages (i.e. for each vowel of a given set there is a counterpart that would receive identical feature specifications if it weren't for that vowel). On the other hand, in languages like Spanish there is absolutely no evidence whatsoever for this kind of subgrouping. Rather, all vowels are approximately equidistant and can therefore be sufficiently well described in terms of height and backness specifications. The case of English, however, is rather more complex. Since this issue is essential to the question of the role of front vowel change in NZE, I will here postpone the discussion to the end of this chapter.

Finally, we note that in a number of languages which can be argued to have the above type of subsystem differentiation, this is to some extent reflected in spelling: Swedish, by and large, uses the same grapheme for corresponding pairs of both sets (e.g. 'kall' with /a/ and 'tal' with /ɑ:/). Similarly, German uses the same symbol for corresponding members, with the addition of a orthographical dichotomiser (in most cases at least, cf. 'Mitte' /mɪtə/ vs. 'Miete' /mi:tə/, 'All' /ɛl/ vs. 'Aal' /a:l/, but cf. 'weg' /vɛk/ vs. 'Weg' /ve:k/). German spelling is therefore rather similar methodologically to the Trager & Smith approach in that in both cases, the set 1 vowels (i.e. the 'short', or 'simple' ones) are regarded as primitives, which combine with something else. The difference is of course that the combined complex structures are explicitly stated to reflect structural complexity in Trager-Smith, whereas it is unknown whether those who orthographised German entertained similar ideas.

We can finish this section by concluding that although there seems to be evidence in favour of the existence of subsystems in vowel inventories both in general and in English, different ways of assessing the reality and the nature of such subsystems do to some extent yield different results. The next section will add to this discussion a diachronic perspective, i.e. take up the claim that different subsystems exhibit differential pathways of change.

### 7.3 Change within/across subsystems

Labov (1994) emphasises that by and large, vowel movements are constrained in such a way that

- (1) In chain shifts, long vowels rise (p. 116).
- (2) In chain shifts, short vowels fall.
- (3) In chain shifts, back vowels move to the front.

These are later (p. 176, 200) restated as:

- (1)' In chain shifts, tense nuclei rise along a peripheral track
- (2)' In chain shifts, lax nuclei fall along a nonperipheral track
- (3)' In chain shifts, tense vowels move to the front along peripheral paths, and lax vowels move to the back along nonperipheral paths.

The restated principles to some extent dissociate the vowel type from its location in articulatory space and introduce the concept of peripherality. In addition, the restatement of 'long' and 'short' vowels into 'tense nuclei' and 'lax nuclei' relates the monophthong system to the diphthong system.

This terminology requires some illustration. The type of vowel system which is expected to show these interrelations can be abstractly sketched as shown in figure 7.1 below.

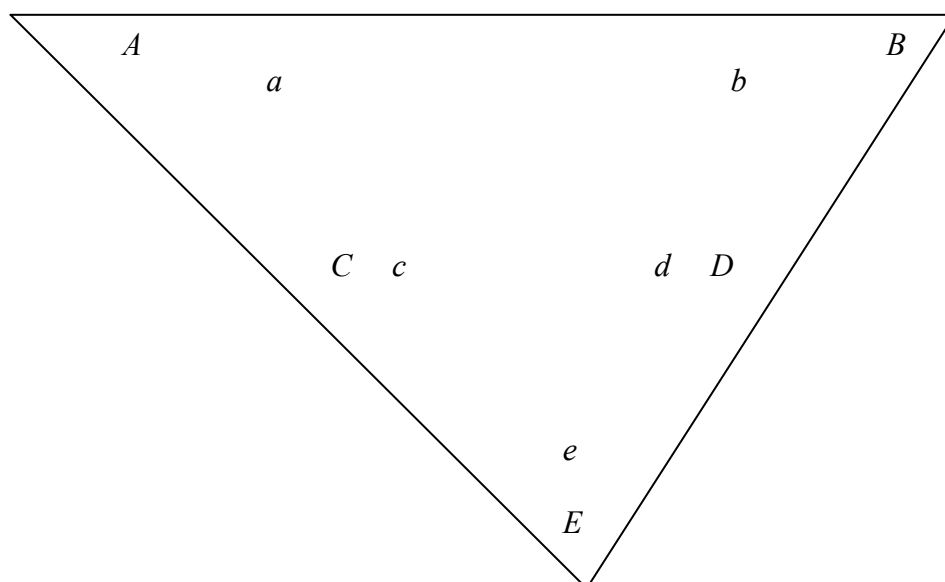


Figure 7.1 – Hypothetical vowel space comprising 2 subsets and 3 heights.

Figure 7.1 makes a number of statements regarding the relational properties of a tense/lax vowel space. First of all, it shows that there are two sets of vowels (/A B C D E/ and /a b c d e/. I will refer to these as {V} and {v}, respectively). The main property of the vowel space outlined by {V} compared to {v} is that it extends further in all articulatory dimensions. That is, the frontest, the lowest and the backest vowels of that system is a member of the {V} set. The {v} set, on the other hand, is a miniature image of the {V} set in that it shows the same types and numbers of oppositions as the {V} set, although the overall space available is more restricted in all dimensions. In addition, the usage of the same letters in both sets implies that there is some relation between *A* and *a* and *B* and *b*, for example. This relation is best illustrated in structural terms: If we look at each of the two subsystems in isolation, we would be able to define each of its members on the basis of a fairly straightforward feature specification (where both *A* and *a* would be [+front] and [+high], for example). This, however, only works if we distinguish between the two sets, since if we fail to do so, different phonemes would receive identical specifications.<sup>47</sup> Adjacency of two vowels in the

<sup>47</sup> Although this is not a paper on phonology, this point deserves a brief digression. Capturing a distinction such as holds between *A* and *a* in figure 1 is somewhat challenging for a feature based view of vowel systems. We would have to assign a further feature other than [+/- high/back] to each member of such a system in order to keep pairs like *A* and *a* distinct. Although a feature such as [+/- long] or [+/- tense] might be used, this would still miss the insight that the same property is used in distinguishing *A* from *a* and from *B*, i.e. quality. Even if we admitted that type of specification, this would result in a specification of *A* as [+high, +front, +long]

overall system is therefore misleading without recourse to the two subsystems (For example, if we had given each vowel of the overall space a discrete symbol, and if vowel shifts involve adjacent segments, we would expect a closer relationship between *C* and *a* than between *C* and *A*). In addition, these vowels are located on ‘tracks’ (p. 177), which means that if they undergo gradual quality change over time (i.e. ‘neogrammarian’ sound change, cf. Kiparsky 1995, Labov 1981, and the classical treatment in Paul 1888) they only do so along trajectories restricted to the members of each set. The ‘two-track’ space can be sketched as in figure 6.2:

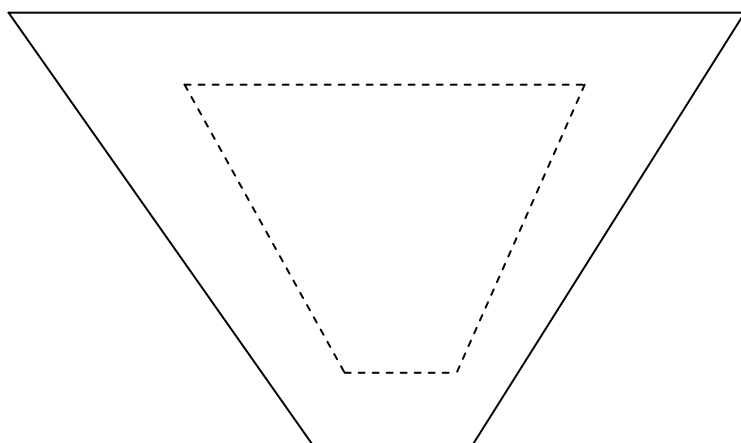


Figure 7.2 – A two-track model of vowel space. The solid line represents the peripheral track, the dashed line represents the non-peripheral track.

The two lines of figure 7.2 represent the two tracks, whereby the solid line corresponds to the peripheral track and the dashed line the nonperipheral track. If we apply this to the idealised vowel system in figure 7.2, we would assign the *V* set to the peripheral track and the *v* vowels to the non-peripheral track as shown in figure 7.3.

---

compared to [+high, +front, -long] for *a*. Although this seems like a valid approach as long as we recognise [+/- long] as an abstract feature, the validity of such a specification rests on its interpretation. That is, do specifications such as these imply that *A* and *a* differ in one abstract feature [+/- long], but are otherwise the same (i.e. both are the same in height/frontness)? Or does it mean that ‘within a given subsystem defined by [+/- long], [+/- high/front] imply a true quality distinction, whereas across subsystems they do not’? It is only on the latter reading where the specification given above is internally consistent. If we give up binary specification in favour of assigning polynary height/backness values to each vowel in the system, we would lose the insight that there are subsystems in the first place. For example, we would specify *A* as [1high/1front], *a* as [2high/2front] and so on. From a purely phonological point of view, this would be viable. However, if we take into consideration the diachronic evidence marshalled in Labov, the interaction patterns between vowels within subsystems become rather mysterious. We would also lack a way of expressing the phonotactic differences between {*V*} and {*v*} as outlined above.

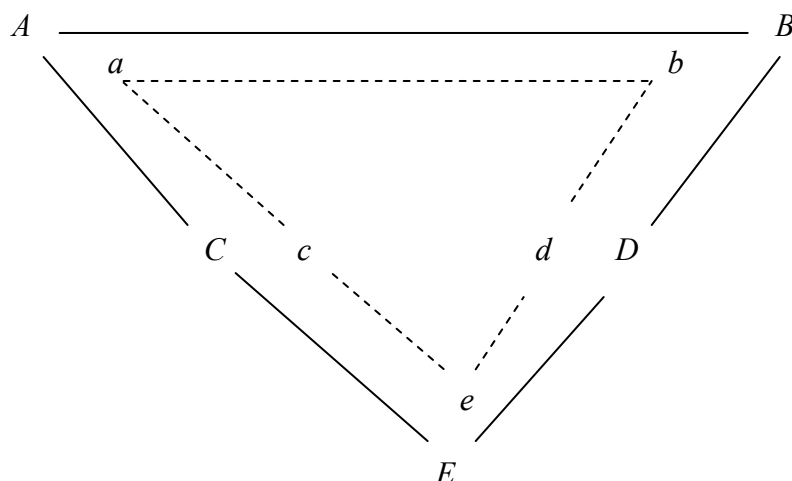


Figure 7.3 – A diagram of a hypothetical tense/lax vowel space, where tense vowels are located on the peripheral track (solid line), lax vowels on the non-peripheral track (dashed line).

If we accept this system, we can redefine adjacency as spatial proximity on tracks. That is, if we assume that a) vowels move and b) they somehow interact (merge with or push out/pull in) with adjacent vowels (and only adjacent vowels), the two-track view resolves the paradox of why, for example, *c* should interact with its spatially adjacent high counterpart *a*, whereas *C* interacts with a vowel one height slot further up (*A*).

The movements of vowels on these tracks are constrained by the principles given above. In this particular case, this implies that *a*, *b*, *c* and *d* would fall along their track, whereas *C*, *D* and *E* would rise. In addition, *A* and *B* would break and develop into diphthongs with a lax nucleus and an offglide which is similar in quality to the former monophthong. Since the principles apply to diphthongal nuclei as well, we would then expect the nuclei of the diphthongs which arose from the breaking of *A/B* to fall. In addition, principle 3 states that *B* and *D* might also become fronted, whereas *c* might be backed. Of course, if we combine the view of phonological space as divided up into tracks with the unidirectionality of vowel change on each of the two tracks, we would eventually expect a two-vowel system with a low lax vowel and one or two high vowels. Since this is obviously not the case, we need one or more mechanisms which allow for some degree of interaction across subsystems. Labov attributes change across subsystems to two ‘exit principles’ (Labov 1994: 280-81), stated as

Exit principle 1: In chain shifting, low nonperipheral vowels become peripheral.

Exit principle 2: In chain shifting, one of two high peripheral morae becomes non-peripheral.

The former of these two principles relates to lengthening/peripheralisation of lax low vowels as exemplified by the lengthening of /a/ in Swedish (cf. Labov (1994: 281), Benediktsson (1970)). The broad-A rule of many varieties of English would be a similar case, and Labov notes that ‘most of these cases [...] involve a great deal of lexical irregularity’ (p 281, footnote 4), which implies that they may not be good examples of stepwise neogrammarian sound change in the first place, but rather operate at a ‘higher level of abstraction’. We will come back to this point later on. The second exit principle relates to the breaking of high vowels as sketched above. If a high monophthong (i.e. bimoraic tense vowel) moves out of its position, it does so by developing either a lax nucleus and a tense offglide, or a tense nucleus and a lax offglide. The former process is exemplified by the breaking of ME /i/ to /ɪi/ > /ai/ during the GVS, and the recent diphthongisation of /i/ to /ɪi/ in the Southern Hemisphere Englishes. Once a high monophthong assumes this type of diphthongal structure, its further development is as expected under the general principles, i.e. the lax nucleus lowers. Once the nucleus has reached the most open position, there occurs a peripherality change-over between the nucleus and the offglide, which implies that further movement of the nucleus adheres to principle 1 (i.e. tense nuclei rise along peripheral paths). This is exemplified by the raising of the nuclei of /au/ and /ai/ in the Southern Hemisphere Englishes.

Finally, Labov adds a further principle which governs the (re)assignment of vowels onto tracks as a function of their spatial relation to other vowels (The ‘redefinition principle’, p 285).

*Peripherality is defined relative to the vowel system as a whole.*

This implies that peripherality is not an absolute position of some element in a vowel system, but is meaningful only if a given vowel can be paired with another vowel of opposite peripherality. This particular principle will later on be shown to be of some relevance in the context of the short front vowel shift. It also follows from this that for a vowel to change

peripherality, it does not necessarily have to move. Rather, it may be reassigned to another track solely on the basis of developments that affect other vowels. In the idealised vowel space above, this might come about by merger of *C* with *A* (on *A*), which would leave *c* in a position where it does not have a peripheral counterpart. It would therefore be indefinable for peripherality. If, in addition, some other vowel *x* moved to a position between *c* and *d*, *c* and *x* would form a new pair of which *c* would be the peripheral member.

Having outlined the basic mechanisms of Labov's theory of vowel movement and vowel space infrastructure on the basis of an idealised vowel space, we can now apply this model to the 'real' vowel space of English. Herein lies a problem: Whereas most other Germanic languages have rather symmetrical vowel systems, where each member of one track can be paired with a member of the other track, the vowel system of most Standard varieties of English is rather asymmetrical in that, whichever way we assign vowels to tracks, there will always remain at least one slot where peripheral to non-peripheral pairing is not possible. Table one illustrates this in comparison with Standard German:

German<sup>48</sup>

	front		central		back	
	p(eripheral)	non-p	p	non-p	p	non-p
high	i	ɪ			u	ʊ
high-mid	e				o	
low-mid		ɛ				ɔ
low			a	ɐ		

English<sup>49</sup>

	front		central		back	
	p	non-p	p	non-p	p	non-p
high	i	ɪ			u	ʊ
high-mid						
low-mid		ɛ (?)			ɔ	
low		æ (?)	a	ɐ		ɒ

<sup>48</sup> Assuming merger of /e:/ and /ɛ:/ on the higher vowel, which seems to be the standard in current spoken German. For the sake of clarity, this table excludes front rounded vowels, which pattern symmetrically with their unrounded counterparts.

<sup>49</sup> Excluding NURSE.

As for the German vowels, the matter is straightforward: Each peripheral vowel in the system has a non-peripheral counterpart. The phonetic oppositions are uniform across the system in that the peripheral vowels are longer and located further toward the outer envelope of the vowel system than their non-peripheral counterparts. In addition, the principles of vowel shifting as outlined above seem to hold in that the Middle High German vowel shift (Pribsch and Collinson (1958), discussed in Labov (1994: 124-125)) affected only the peripheral vowels in a fashion similar to the Great Vowel Shift of English. The interaction of vowels within tracks is furthermore illustrated by two mergers: lax /e/ and /ɛ/ merged on /ɛ/ (cf. the homophony of *schlechter* ‘worse’ and *Schlächter* ‘slaughterer, i.e. mass murderer’, which are both /ʃlɛçtɐ/), whereas their tense counterpart recently merged on the higher vowel (so that for many speakers *sehen* ‘to see’ and *säen* ‘to sow’ are homophonous: /ze:n/). In addition to suggesting the existence of tracks, these two mergers also illustrate the opposing directions of movement on the tracks.

With respect to the English system, the matter is more complicated due to the uncertain status of /æ/, from which follows a similar degree of uncertainty with regard to /ɛ/<sup>50</sup>. If we look at position in articulatory space alone, it might seem valid to impose symmetry onto these two vowels by assigning /ɛ/ to the peripheral track and /æ/ to the nonperipheral track. We would then mirror both the set-up of /o/ and /ɒ/ in the front region as well as the front vowel set-up of modern German. This analysis, however, is at odds with what we know about these two vowels. Both their phonology as well as their history shows that whatever they are in terms of peripherality, they are the same since they share a number of properties:

- Both /æ/ and /ɛ/ are phonotactically restricted in that they cannot occur in stressed unchecked syllables.
- Both vowels derive from Middle English short vowels.
- At least in the data reviewed in this thesis, their duration is similar (cf. chapter 3)

---

<sup>50</sup> Note that the overall pattern is somewhat abstracted from phonetic reality as it assumes that START and STRUT are central vowels, which is arguably not the case in RP and other varieties, where at least START would be a back vowel. Similarly, it is questionable whether GOOSE is best regarded as a back vowel. However, these points refer to the phonetic realisation of these vowels and do not directly bear upon the question of subsystemic patterning. That is, each of these vowels can be paired with another vowel of opposing peripherality and shows the expected relationships. For example, whether GOOSE is analysed as a back vowel or a central vowel, this does not have any impact onto the status of GOOSE as a tense/peripheral vowel since it shows both the phonotactic as well as the diachronic properties of tense vowels.



- They interact with each other in chain shifts, which would be unexpected if they were located in different subsystems.

Some of the above points are contestable on the basis of data from American English. First of all, Labov shows that /æ/ undergoes tensing in certain phonemic environments in a number of varieties of American English whereas no such process is reported for /ɛ/. In addition, it has been shown that /æ/ is considerably longer than other originally short vowels in American English (cf. the references given in chapter 3). However, even if we allow /æ/ to be analysed as a tense vowel, we would still be faced with an asymmetry between front and back vowels in that for the front vowels, it would be the lower vowel /æ/ which is located on the peripheral track, whereas in the back vowels it is the higher vowel (/o/). Before discussing the implications of this asymmetry, we will briefly sum up the historical developments in the English vowel system that led to this peculiar situation. Two processes are of importance in this respect: The Early Modern English merger of ME /e:/ and /ɛ:/ as well as the lowering of ME /u/ to /a /, which subsequently pushed out /a/ to a low front position (cf. Trudgill 1986). The former process had two structural consequences: First of all, it removed one peripheral vowel from the overall monophthong system. In addition, the merged vowel rose to high front position. This created two asymmetries, one within the subsystem of peripheral vowels and another one across subsystems. The former relates to the fact that the post merger/raising system has a peripheral mid back vowel without a counterpart in the front peripheral series. In addition, the process creates a numerical asymmetry across systems by decreasing the number of peripheral vowels with no corresponding change in the non-peripheral system.

The second process was a split of ME /u/ into a high back vowel (FOOT) and a back-central low vowel (STRUT), which in turn brought about a fronting of the former /a/ vowel to [æ]. Again, this process introduces two further asymmetries in that it adds a further vowel to the system of lax vowels and therefore introduces a front-back asymmetry within the non-peripheral track (since one non-high back vowel now patterns with two non-high front vowels on the nonperipheral track). The process can be diagrammed as follows:

## ME front vowels 1

	long	short <sup>51</sup>
high	i:	i
mid-high	e:	
mid-low	ɛ:	ɛ
low		

This is essentially the situation throughout most of the Middle English period. Although there exists some asymmetry between subsystems in that the long vowels have one additional member, this was reflected in the back vowel series as well (where Modern English LOT = /ɔ/, GOOSE = /o:/, GOAT = /ɔ:/). Each member of the ‘short’ set has a long counterpart, although the reverse does not hold. However, a gap in the non-peripheral space is rather more plausible because if we assume that in a period of stability vowels tend to maintain equal ‘margins of security’ (Martinet 1955) around their mean, the peripheral vowel space can accommodate more vowels than the nonperipheral space due to the fact that the envelope it encompasses is larger. After the merger of ME /e:/ and /ɛ:/ and the subsequent raising of the merged phoneme to /i:/ in the GVS, the long vowel system was restricted to one member in the front series.

---

<sup>51</sup> The ‘short’ vs. ‘long’ terminology may be valid in the case of Middle English since it is not clear when the breaking up of the older system which relied purely on length in pairs like /i/ vs. /i:/ actually occurred. Luick (1914) dates this back to Early Middle English on the grounds that in the process of Open syllable Lengthening, some OE high short vowels became long midvowels (cf. Old English *wicu* ‘week’, which after OS� had a mid vowel that was subsequently raised to high position in the GVS), which seems rather more plausible if we assume that they were dissimilar in height and backness to the long vowels (that is, they were both lower and more central than their long counterparts) and may therefore have been associated with a long /e:/ rather than /i:/ after lengthening. Lass (1997), on the other hand, prefers a later date (as late as the 18<sup>th</sup> century) on the grounds that there is no evidence whatsoever in the otherwise rather accurate reports by the early orthoepists, who Lass otherwise assumes to be rather accurate observers of articulatory patterns.

## Early Modern English (pre-/u/ split)

	long	short
high	i:	i
mid-high		
mid-low		ɛ
low		

The second process then adds a further member to the short series of front vowels, which results in the following modern system.

	long	short
high	i:	ɪ
mid-high		
mid-low		ɛ
low		æ

This system is unparalleled in other Germanic languages in that

- No other Germanic language has more lax vowels than tense vowels.
- No other Germanic language has a vacated peripheral track between /i:/ and /a:/.
- No other Germanic language has more oppositions in the series of short front vowels than in the short back vowels.
- No other Germanic language has ‘orphaned’ lax front vowels with no tense counterparts.

If we furthermore take into account the dynamical aspects of this systems from the early Modern English period onwards, we can add that

- No other Germanic language (including the predecessors of Modern English) shows a comparable amount of variability in space and time with regard to the realisation of /æ/ and /ɛ/.

The last statement sums up a number of facts from both the realisation of /æ/ and /ɛ/ across different varieties of contemporary English as well as their development in these varieties. These can be summed up as follows:

- In Southern British English, /æ/ can be either a low-mid front vowel (as in Cockney), a low front vowel (as in ‘classical’ RP) or a low central vowel (as in ‘innovative’ RP and (or?) ‘Estuary English’. This implies that if we assume that they all derive from a vowel of similar quality, we would infer either
  - (1) Raising of ME /a/ to [ɛ] ~ [e] and a subsequent lowering to [a]. This implies that, of the varieties mentioned above, Cockney would have the most conservative realisation of /æ/, whereas innovative RP would have the most innovative one.
  - (2) Raising of ME /a/ to [ɛ], with no subsequent lowering. This implies that Cockney would have the most innovative vowel, and modern RP the most conservative one.
  - (3) Raising of ME /a/ to [æ] and subsequent rising and lowering in different varieties (where Cockney and modern RP would then be equally innovative, whereas traditional RP would have the conservative vowel)<sup>52</sup>.
- In most varieties of American English, /æ/ is a low front vowel [æ]. In addition, /æ/ is considerably longer in GenAm than in RP or the Southern Hemisphere varieties of English. Extensive research by Labov (summed up in Labov (1994/2001) has shown that in some varieties, the /æ/ phoneme shows extensive allophonic variation with values ranging from [æ] to [ɪ] ~ [ɪə] depending on either phonological context or lexical identity. Labov goes on to show that in the varieties which have this type of

---

<sup>52</sup> This matter is not entirely settled. For a discussion see Trudgill (2004: chapter 2), who argues for a scenario depicted in (1) rather than (3), the latter being the more widely accepted view (cf. Bauer (1979/1992)), Lass and Wright (1985).

variation in /æ/, [æ] seems to be the original value and the subsequent developments can be described as *tensing* and *lengthening* of a number of allophones of /æ/.

- In the Southern Hemisphere varieties, /æ/ is generally higher than [æ]. In SAE and AusE, the most common realisation is [ɛ], although innovative speakers of AusE show lower values (cf. Cox (1996) and Bernard (1970)). The NZE /æ/ phoneme is realised as an even closer vowel.

Given the above generalisations regarding the realisation of /æ/ in Modern English and the historical fact that /æ/ derives from a short central vowel /a/ in Middle English, the minimal number of sound changes affecting /æ/ which we have to posit would be one process of raising (ME /a/ to [e], where each of the above varieties has stopped at different stages of the raising) as well as at least one process of lowering (in innovative RP and possibly innovative AusE). Although the actual picture is much more complex<sup>53</sup>, even this minimum seems to contradict Labov's theory that vowels of the same type behave uniformly over time.

A similar case can be made for /ɛ/, which shows at least lowering in some varieties of American English as well as raising in the Southern Hemisphere varieties (to a value approximating [e] in AusE and [ɪ] in NZE). A further example of the non-uniformity of a vowel phoneme with regard to its trajectory over time involves the realisation of /ɪ/ in two closely related varieties of English, namely AusE and NZE. Both the Australian and the NZ realisation have been shown to have developed from the same original vowel [ɪ] (at least in the case of NZE, the matter is now undisputed. Recent research on earlier forms of AusE points in the same direction. P. Trudgill and E. Gordon, p.c.), which implies that in AusE, /ɪ/ has been raised and fronted, whereas in NZE, it has centralised and lowered.

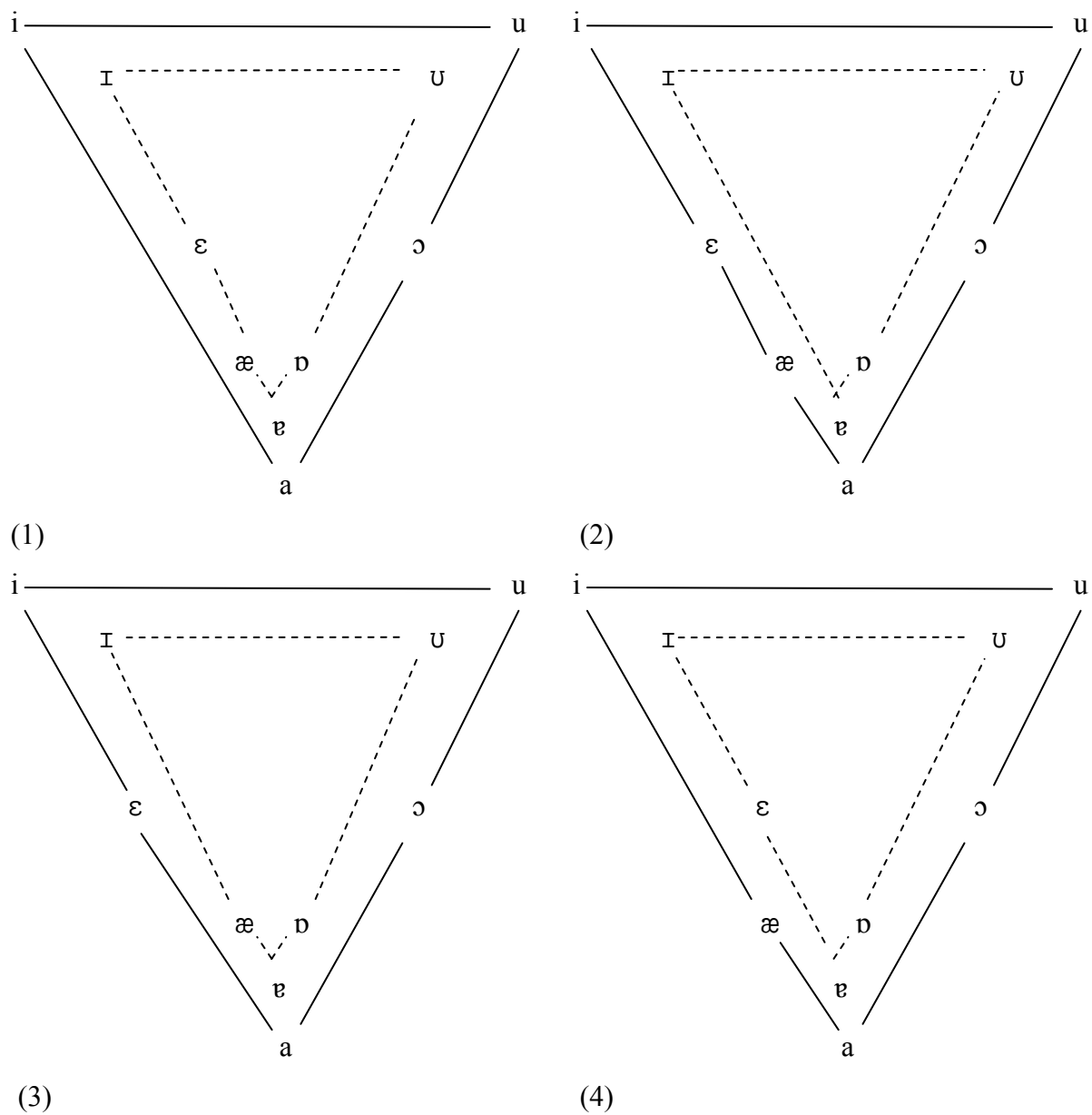
We are therefore faced with three vowels which seem to violate Labov's theory of directional vowel change, irrespective of which value is assumed to be the original one in each case. In the following two sections, I will argue that the variability of the non-high front vowels is probably due to different interpretations of peripherality in different varieties of English.

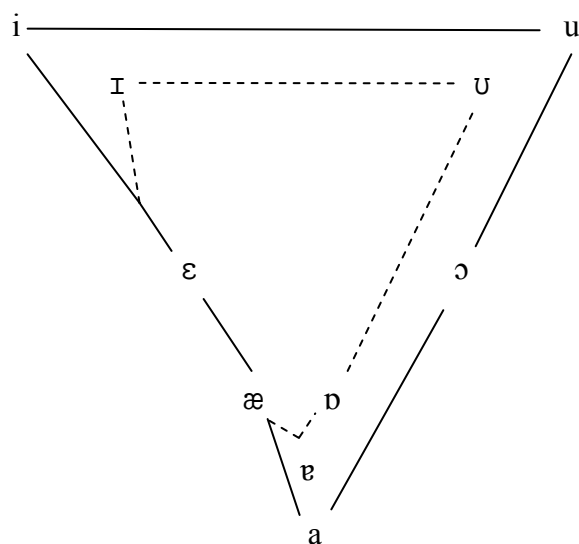
---

<sup>53</sup> Additional complexity arises since it is not clear to what extent entities such as 'RP' and 'Estuary English' are truly different varieties in a geographical sense, or whether they represent mainly register differences, or a mixture of both. In addition, there now seems to be some evidence for retrograde movements of /æ/ in Australian English (cf. Cox 1996 and Bernard 1970).

## 7.4 Front vowels on tracks

If we recognise the existence of two distinct tracks within the vowel space of English, and if we furthermore assume that tense vowels are located on the peripheral track, whereas lax vowels are located on the non-peripheral track, and if we take into consideration the uncertain status of /æ/ and /ɛ/, any one the following possible patterns of allocating the vowels of English onto tracks may seem valid:





(5)

The vowel space as outlined in (1) assumes that both /æ/ and /ɛ/ are non-peripheral vowels and therefore pattern with /i/, /a/, /ɔ/ and /u/. This would capture the facts that a) /æ/ and /ɛ/ are similar in length to the lax vowels and b) interact with the lax vowels in dynamic sound change. In addition, it associates the vowels' location in articulatory space with their phonotactic properties (i.e. no non-peripheral vowel can occur in free stressed syllables). However, this representation fails to account for why both /æ/ and /ɛ/ can move in different directions on their tracks, whereas other vowels do not. More specifically, this localisation would predict uniform lowering of both /æ/ and /ɛ/, which is not the case.

Similarly, the patterning in (2) (where both /æ/ and /ɛ/ are located on the peripheral track) would make the wrong predictions in that it would only allow for upward movements. In addition, it would dissociate the phonotactic properties of vowels from their phonological subgrouping (since 2 vowels which cannot occur in free stressed syllables form part of the set which otherwise can occur in that environment).

The representations in (3) and (4) both locate the two non-high front vowels on different tracks. In (3), /ɛ/ is represented as a peripheral vowel and /æ/ as a non-peripheral vowel. This misses anything we know about the two vowels, in that it fails to predict that the two vowels have the same phonotactic properties, that they interact with each other in vowel changes, and that if any of the two vowels has phonetic properties more similar to the

peripheral vowels, it would be /æ/ (at least in American English). It also fails to account for the different behaviour of each of the two vowels across varieties.

(4) is the mirror image of (3) in that it locates one vowel (/æ/) on the peripheral track and the other one on the non-peripheral track. Although this seems like an equally unlikely pattern, it might indeed be appropriate for those varieties of American English which raise /æ/ (or rather, certain allophones thereof) to a close position without any reactionary movement of /ɛ/. However, it fails to account for the lowering of /æ/ and the raising of /ɛ/ in varieties where these processes have been found.

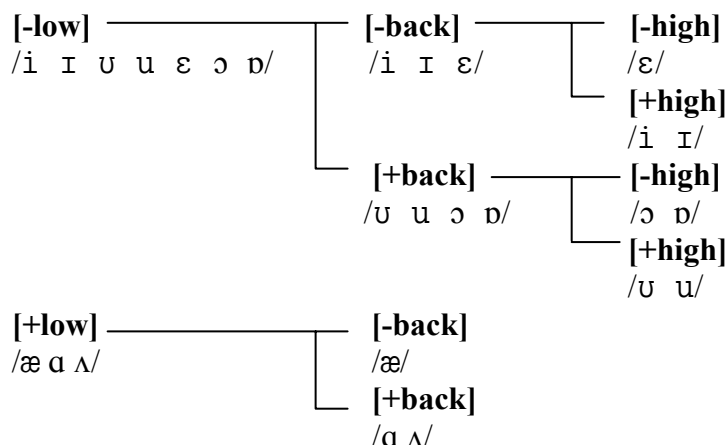
Finally, the representation in (5) states that the two vowels /æ/ and /ɛ/ are not definable *per se* in terms of their location on articulatory tracks. As such, this would allow each of the two vowels to be (re)interpreted as peripheral or nonperipheral vowels at any time and would therefore account for the differences in their behaviour. However, this seems like an ad hoc way of forcing a preferred pattern upon the real world and therefore requires some justification. The idea was entertained by Labov (1994) and is essentially summed up by the principle that ‘peripherality can only be defined by the vowel system as a whole’. This means that peripherality requires one or more reference points in articulatory space. If we recall the abstract tense/lax vowel space as outlined in figure 2 above, we see that each vowel of the set {V} was in some meaningful sense paired with one member of {v}, be it etymologically or phonologically. Therefore, each vowel in that system has a reference vowel of opposite peripherality. This follows directly from the subdivision of a formerly purely quantitative vowel space into peripheral and nonperipheral sets. If we translate this into the vowel system of Modern English, the vowels /ɪ ʌ ɒ ʊ/ all have a counterpart in the set of tense vowels in both etymological and phonological terms.<sup>54</sup> For historical reasons discussed above, this is not the case for /æ/ and /ɛ/. If we put this in phonological terms and recognise [+/- peripheral] as a feature, /æ/ and /ɛ/ are the only vowels in the overall system which do not have a

---

<sup>54</sup> The etymological pairing, however, does of course not imply that the modern /ɪ/ - /i/ pair is historically ‘the same’ as the Middle English /i:/ - /i/ pair since the GVS raised the long vowels by one slot. It rather means that the modern lax vowels /ɪ ʌ ɒ ʊ/ have a tense counterpart that originated as a long vowel in ME. In addition, the relation between /ɪ ʌ ɒ ʊ/ and their original counterparts is in principle still derivable by rule (cf. Chomsky and Halle (1968)), since the mapping is straightforward. This is decidedly not the case for /æ/ and /ɛ/, who either do not have an etymological ME counterpart (/æ/) or where this counterpart has merged with another phoneme (/ɛ/). Cf. the discussion in 6.3 above.



counterpart which shares all other features except this one. A summary of basic feature specifications is given below.



It seems clear that whereas for all other vowels, a feature such as [+/- peripheral] would be necessary for each of them to receive a unique feature specification, this is not the case for /æ/ and /ɛ/.

We therefore have both historical as well as phonological reasons for assuming that these two vowels do in fact not pattern with other vowels in the same way as any other vowel does. In addition, we have seen in chapter 3 above that in the sample analysed here, /æ/ and /ɛ/ show a number of phonetic properties that are to some degree more typical of tense vowels, in that they are longer than some lax vowel (/ɪ/) and more susceptible to allophonic shortening in applicable environments.

This analysis accounts for a range of phenomena involving /æ/ and /ɛ/ across different varieties, namely that

- Either of them can undergo raising or lowering.
- Although it is often the case that they interact (e.g. both undergo raising successively), this is not necessary (as in the case of the Northern Cities Shift). For the latter to happen, it is required that only one vowel (or a number of allophones thereof) undergoes relocation to another track<sup>55</sup>.

<sup>55</sup> Although I am not aware of whether this has been addressed in Labov's work, it would be interesting to see whether (or to which extent) those allophones of /æ/ that eventually underwent raising in the Northern Cities

## 7.5 Conclusion

The aim of this chapter was to provide a general background discussion of how vowel spaces in general, and the vowel space of English in particular, are structured. It was shown that one might arrive at different views depending on whether one prefers phonetic, historical, or phonological criteria. I have also discussed the role of the short front vowels within the overall vowel system of English. Within the model advocated by Labov, these seem to behave in an aberrant way in certain varieties of English, most notably the Southern Hemisphere varieties. I have attempted to reconcile the developments in the Southern Hemisphere short front vowels with Labov's model by proposing to relax of some of the primitives of the original framework. More specifically, I argued that the concepts of 'track' and 'peripherality' are useful only if they are applied locally rather than globally, which is to say that they hold only in regions of the vowel space where instantiations of appropriate vowels exist in the first place. Given the historical contingencies of vowel shifts in English as outlined in section 7.3, it was proposed that English front midvowels constitute a case where a model that requires tracks and vowel pairings of opposite peripherality fails. Rather, it was concluded that this part of the vowel space is unstructured in terms of tracks and peripherality, and that the status of the vowels in that region is *indeterminate* with respect to these categories.

Now, if a good case can be made that some vowels show this indeterminacy with respect to their location on tracks, it follows that the concept of 'tracks' is not universally required in order to make sense of the historical developments of complex vowel systems. In the following chapter, I will go one step further and argue that this concept could be abandoned altogether, and that chain shifts can be understood by resorting to one fundamental functional mechanism, the increase of distance between vowel distributions. I will propose a model which accounts for the local interrelationships between vowels in a small-scale chain shift (such as the short front vowel shift of New Zealand English) by resorting to only a small number of basic principles in the organisation of vowel systems of a certain kind and their subsequent development. I will show that most of what seem to be

---

Shift) are phonetically different from those allophones that remain [æ] in the speech of earlier speakers who have not yet split /æ/. The model proposed in chapter 8 would of course predict such differences, e.g. in vowel duration.

disparate processes (such as raising, lowering, and centralisation) follow from the basic distributional properties of a tense/lax vowel system.

## 8.

**Chain shifts in multidimensional vowel spaces – an evolutionary approach**

## 8.1 Introduction

In this chapter, I will propose a functional model of vowel shifts which is in principle capable of making concrete predictions regarding the pathways of such shifts. The mechanism proposed here is similar to those established in evolutionary biology, namely selection of a subset of variants of a given archetype (here: the real-world instantiations of a vowel phonemes). I will make a case for ‘fitness’ as an important parameter in the evaluation of vowel tokens. The chapter is structured as follows: section 8.2 presents an outline of the basic mechanism and exemplifies this mechanism on the basis of hypothetical vowel distributions. In section 8.3, the model will be expanded by arguing for the importance of phonetic dimensions other than F1 and F2 in assigning fitness to vowel utterances. Both hypothetical vowel spaces as well as real-world data analysed in chapter 3 will be discussed. Section 8.4 will return to the main topic of chapter 7 by claiming that under the model proposed here, the concept of ‘track’ can be shown to be redundant, i.e. an epiphenomenon of a rather more simple principle governing chain shifts, namely the *interaction/adjacency principle*. Data from the intermediate sample will be used in order to substantiate the implications of this concept. A small number of related phenomena will also be discussed. Finally, some important caveats will be voiced in section 8.5.

## 8.2 A mechanistic model of push shifts.

I propose that if a vowel A moves and thereby approximates a neighbouring vowel B, the following mechanisms/principles hold:

- Maintain distinct distributions.
- In order to do this, generalise a new mean of the overall distribution of B over a range of allophones of B that

- are furthest away from A

*or:*

- lie outside the trajectory of movement of A.

*and*<sup>56</sup>:

- encroach least upon another neighbouring vowel C.

If we accept this line of reasoning and take it as a given phonemes are by default kept apart (we will discuss problems with this working hypothesis below), the particular processes that we are concerned with here follow by adhering to a linguistic version of Darwinian evolution. Of course, this image is by no means new and has been discussed at length with respect to both its merits and its problems<sup>57</sup>. The major problem in equating biological evolution with linguistic evolution has been to find an adequate correlate of biological ‘fitness’ in language, as it remains to be shown that the evolution of languages is anything but neutral with reference to their functional suitability. This problem is stated in Labov (2001:14) as the ‘Darwinian Paradox’:

#### *The Darwinian Paradox*

The evolution of species and the evolution of languages are identical in form, although their fundamental causes are completely different.

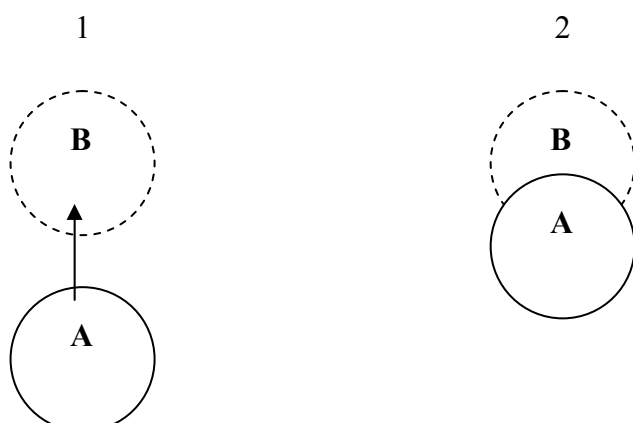
---

<sup>56</sup> The *and* has scope over both preceding principles.

<sup>57</sup> The insight that language change and organic evolution show similar mechanisms goes back to Darwin himself (1871). However, serious evolutionary reasoning/modelling as applied to language change is rather recent (cf. Ohala (1989), Lindblom (1984/1998), de Boer (2000/2001), Blevins (2004), Croft (2000), Keller, Lass (1997)).

This is to say that the fundamental functional driving force behind evolution (i.e. natural selection of individual phenotypes) can not be assumed to have much force in language change, since there is no evidence that languages become ‘fitter’ as they change. However, if we restrict this image to the processes of interrelated chain shifts, I do think that a case can be made for ‘fitness’. On this account, the fitness of some variant is equivalent to it being an unambiguous instance of the phoneme it encodes. In other words, if some variant of vowel phoneme X is prone to be confused with instances of some other phoneme Y, it would be of lesser fitness than some other variant of X. In more concrete term, the fitness of some variant in vowel space is therefore a function of its distance from the distributions of other vowel phonemes. Taking the evolutionary analogy further, the vocalic ‘gene pool’ would be the overall range of variants of any vowel phoneme.

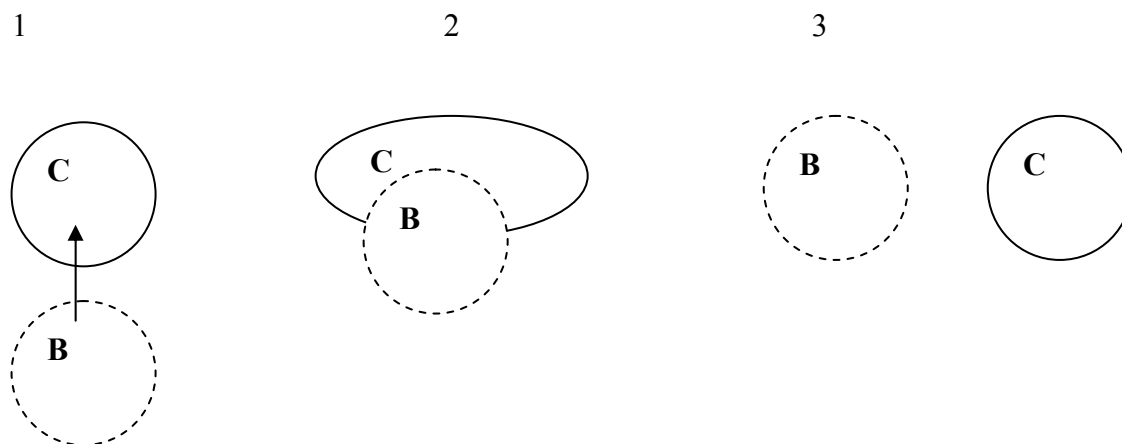
What is more problematic is the notion of **selection** in language change. In biology, selection is an essentially agentless process, i.e. an ‘emergent process’ resulting from the fact that a) there is more offspring than the environment can cater for b) the organism that is less adapted to its environmental conditions has a higher chance of dying before reaching breeding age and therefore c) the genetic make-up of the better adapted organism is ‘selected’ due to its having a higher chance of surviving until it reaches reproductive age. With regard to the transmission and development of vowel systems we would have to axiomatically posit **propensity of not being confused with another vowel** (i.e. a potential manifestation of another vowel) as a correlate of fitness in order to derive a mechanistic evolutionary process from there. For the case at hand, let us assume an upward movement of a low mid-vowel to mid position (i.e. a hypothetical correlate of TRAP raising). This can be schematised as follows, where A is a distribution of a vowel which undergoes upward movement and B is a distribution of another vowel (here a mid-vowel, a correlate of DRESS). Let 1 and 2 be two successive stages of development.



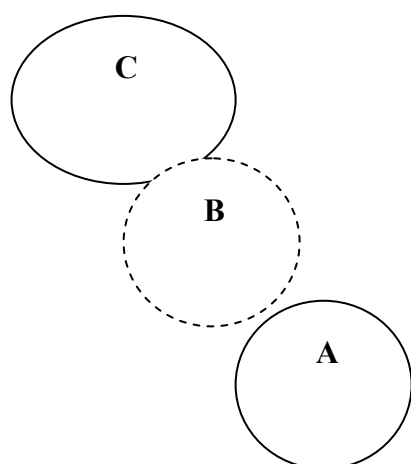
Now given that a learner is exposed to system (2) under the assumption that on some occasions, he wants to signal the difference between a word containing A and another word containing B<sup>58</sup> and given that the two vowels differ only in the height dimension, he would experience a higher success rate by selecting those tokens of B which are located toward the high end of the distribution. This means that success in discrimination between A and B correlates only with the height dimension. As a further step, the learner might generalise the overall distribution of B to what was formerly a group of (high) allophones of B. Therefore, the trajectory of movement of B is predictable once the initial conditions are known (note that the front-back dimension would not help the learner in this scenario, since both A and B have tokens of the same frontness. The learner would not be able to deduce a correlation between frontness and discriminability).

This is fairly straightforward and extends only to a push-chain relation between two vowel distributions along one dimension. In the case of the NZE front vowel shift, however, there is a third phoneme involved (/ɪ/), originally a high front vowel. Again, let us schematise the initial state of affairs between B and C (/ɪ/) in an idealised way; 1/2/3 being successive stages of development.

<sup>58</sup> 'Signalling' here simply means 'intending to associate a given concept with a phonetic form containing a given phoneme', i.e. plain 'speaking', not 'intentional signalling'.



That is, C reacts to the upward movement of B in that it develops an intermediate distribution with fronted and centralised allophones at stage 2 and a central overall distribution at stage 3. It is clear that neither the fronting nor the centralisation are in any way predictable from the state of affairs at stage 1 according to the model outlined above. However, this is probably due to the fact that the initial state (i.e. the inferred ‘unshifted’ set-up) of the three vowels involved is inaccurate (‘overridealised’) and that in most varieties of English, height correlates with frontness in the system of front vowels under question. That this is the case is shown by overall formant frequency plots such as Peterson and Barney’s (1952). A more appropriate schema of English short front vowels of the RP/GenAm type would therefore look rather more as follows:

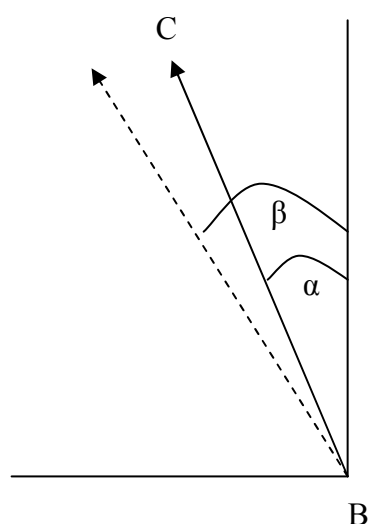




Under this initial model, and assuming movement of A toward the space occupied by B, the option ‘raise B’ would still constitute a viable process of disentangling A and B. Both height and frontness of B would correlate with discriminability between A and B, which implies that a combination of these two processes is not surprising.

The relation of B and C, on the other hand, is rather more complex. As was shown in chapter 3 above, early intermediate speakers show both raised/fronted allophones of /ɪ/ as well as centralised variants (which mirrors the synchronic situation in Australia vis-à-vis New Zealand). As was the case with the first raising, the raising and fronting (of pre-velar allophones of /ɪ/ in the speech of the early Intermediate speakers, and of /ɪ/ as a whole in AusE) makes sense under the present model. The centralisation process, which finally wins out in NZE, is not so easily accounted for. An explanation based on the model proposed here would have to answer two questions: (a) are there allophones of C which can in some sense be regarded as ‘more central’ relative to the most central allophones of B (since further centralisation can only work on existent variation including centralisation) and (b) where do these allophones come from? On the basis of the data from early intermediate speakers it seems as if centralised variants were indeed present from the onset of this period onward. However, I argued above that these speakers already show incipient shift between B and C (i.e. /ɪ/ and /ɛ/), which would of course imply that the centralised variants (i.e. variants of /ɪ/ which are more central than any variant of /ɛ/) are the result of the upward movement of /ɛ/ rather than a precondition, which would make the model circular. In addition, Gordon et al. (2004) show that central variants of /ɪ/ are comparatively rare in the speech of first and second generation New Zealanders. However, if we recall the trajectory of movement of B (‘DRESS’ raising), we note an association of raising and fronting as linked parts of the overall process involving B. This implies that at any point in the process of decreasing discriminability of B and C, those tokens of B which are closest to C in the height dimension will be those that are furthest from the most central allophones of C in the front/back dimension. This in turn correlates centrality in C with discriminability from B. Whether centralisation or fronting wins out eventually seems to be a non-predictable choice between two equally valid options. However, on the basis of the above diagram it may not at all be clear that the central allophones of C are in any sense valid candidates for selection under the

principles discussed so far, at least on the basis of the above diagram. Rather, the degree to which these central variants can be justifiably assumed to be ‘fitter’ than other ones depends on the original configuration between B and C. That is, the lesser the difference between B and C on the horizontal (provided that B is more central on average to begin with), the higher the likelihood of central variants of C being likely candidates for selection. In numerical terms, it would have to be shown that the trajectory of movement of the lower vowel is directed toward the front end of the distribution of the higher vowel. This is not a straightforward point to demonstrate since one would have to overlay the attested trajectory to an initial pattern *post hoc*, since we do not have detailed information on trajectories of movement in earlier (pre-NZ, early NZ) stages of the shift. This can be diagrammed as shown below:



We can think of the initial constellation between the two vowels as the angle between the y-axis and a straight line between the two vowels (the solid arrow). The trajectory of movement of B may be given by connecting the average position of DRESS at two successive stages of development, and calculate the angle to the y-axis. The former angle is represented as  $\alpha$ , the latter as  $\beta$  in the diagram. For the above model to work, one would have to show that  $\beta > \alpha$ . If we look at the data from the intermediate male speakers, this is borne out. For these speakers,  $\alpha \sim 0^\circ$  and  $\beta = 45^\circ$  (of course, the absolute angular values are dependent on the overall scale of the graph from which they are derived, but the relative values, which are relevant in this

context, do not). This was determined by overlaying the trajectory of DRESS raising/fronting over the intermediate male sample (i.e. a line connecting the average DRESS for the early males and the late males) and the initial relation between KIT and DRESS (i.e. a line connecting the early male DRESS and KIT averages). Although data from earlier periods of NZE is shown in Gordon et al. (2004: 109), their graphs do not show successive stages of development in earlier NZE. It is therefore less straightforward to make a similar calculation for their speakers. The data is, however, broken down into male and female speakers, and it is clear that the female speakers show a more advanced setup of short front vowels. We might therefore justifiably treat the female speakers as exemplifying a ‘later’ stage in the development, which would give us a basis for calculating  $\beta$  for MU speakers (where  $\alpha$  would represent the angle between the y-axis and a line connecting the averages of KIT and DRESS from the MU male sample). If we do this, we obtain  $\sim 20^\circ$  for  $\alpha$  and  $\sim 70^\circ$  for  $\beta$ . This suggests that the necessary condition for central tokens to receive a relatively high fitness value, namely that the trajectory of movement of DRESS is toward the front end of the distribution of KIT, is met in both cases.

At any rate the diagrammatical illustration above is only part of the story, since it does not yet take into account the duration dimension (cf. 7.3 below), which plays a role in this regard.

It therefore seems that a mechanistic model that relies on generalising an overall distribution to a range of previously existent allophones allows us to account for the SFV push chain without invoking a *deus ex lingua* who enforces some overall process that makes sense only from a post hoc perspective. What appears as a result of teleological mechanisms (‘clean up your crowded front vowel system by changing it into a system of two front vowels and a central vowel’) emerges automatically as an incidental result of small-scale transmission processes which are much more plausible (reliable discrimination of two phonemes). This can be summed up by the first two mechanisms outlined at the beginning of this section:

- Maintain distinct distributions.
- In order to do this, generalise as a new mean of the overall distribution of B a range of allophones of B that
  - are furthest away from A (governs DRESS fronting/raising and KIT fronting/raising).  
This can also be read as: ‘Allow for the greatest degree of discriminability from A’.

*or:*

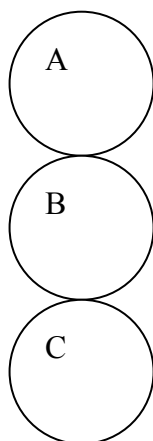
- lie outside the trajectory of movement of A. (governs KIT centralisation)

The relationship between DRESS and TRAP is less easily accounted for. After all, under the mechanisms proposed here, centralisation of DRESS as a reaction to initial TRAP raising would be a valid option. However, neither the intermediate data nor Gordon et al’s allow for clear statements regarding trajectories of movement for TRAP relative to DRESS. Comparing Gordon et al’s male and female sample (p. 109), there is almost no difference between the samples with respect to the position of TRAP. Although the intermediate data shows change in TRAP, it is inconclusive in two respects: First of all, there is a difference between male and female speakers with respect to the TRAP trajectory relative to the DRESS average (cf. figure 3.4). If we infer a trajectory by connecting the averages from the early and the medium stage, the male speakers’ TRAP vowel heads toward the front end of the (early male) DRESS distribution, while that of the female speakers heads toward the central part (of the early female distribution). In addition, the trajectory is not linear in any of the two samples, since late speakers show a more central realisation of TRAP than medium speakers. These points make it impossible to apply a rationale for why DRESS raised/fronted rather than centralised/lowered on the basis of the available data. On the other hand, the more central position of TRAP in the late samples compared to the medium one is intriguing in itself. Although quite possibly due to sampling error, it would follow from the model: As TRAP approaches DRESS (the latter being further front during the initial stage) in the F1 dimension, there would hold a stage where central tokens of TRAP would receive a relatively high fitness value. This would lead to

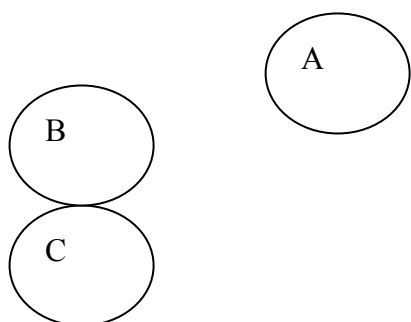
some degree of centralisation in TRAP, which would in turn feed DRESS fronting. With subsequent DRESS raising/fronting and KIT centralising, this ephemeral trend would be reversed, allowing for further fronting of TRAP.

On the whole, the above model represents a reformulation of Martinet's idea of 'margins of security', which in Martinet's approach remains a rather mystical concept, since it is not clear what this 'buffer zone' actually is. On the account proposed here, the margin of security of some vowel represents the vowel space populated by real-world instances of that vowel. Threatening of these margins would be equivalent to token overlap between two adjacent vowels.

I would also maintain that this mechanism is in principle capable of accounting for pull chains as well. More specifically, pull chains come about as the result of the high degree of fitness of variants of some vowel distribution that are close to a 'void' in a vowel space. Once this original distribution moves in that direction, the distance relations to its formerly adjacent distribution are redefined, which then changes the fitness values within that distribution as well. Assume the following adjacent vowel distributions differentiated by height.



Within B, none of the tokens are fitter than any others. However, assume A moves toward the centre of the vowel space:



Once A has progressed in that fashion, the higher tokens of B are better predictors of ‘Bness’ than the lower ones are, and have therefore acquired higher fitness, i.e. propensity of being selected. The movement of A has thus brought about a skew in fitness within the distribution of B, which would then bring about a subsequent movement of B away from C.

This leaves us with the complementary type of gradual sound change, i.e. merger between two phonemes. On the face of it, we have an initial situation which is similar to that between either two of the short front vowels, i.e. an upward movement of some vowel (/ɛə/) into the vowel space occupied by some other vowel (/iə/). The question is why the push-chain type of process we observed in the short front vowel system is not mirrored in the centring diphthongs. I can not offer an alternative account along the lines proposed here, and a phonemic merger may therefore be due to completely different mechanisms such as informative load (Wang 1967, Labov 1994) or symmetry (Martinet 1955, Hawkins 1976, cf. also the following chapter of this thesis), none of which are particularly convincing (this is discussed at length in Lass 1980/1997).

### 8.3 Vowel shifts in multi-dimensional vowel spaces

Up to this point, the discussion has focussed on the arrangement of vowels and vowel distributions in two-dimensional space. However, recall the evidence presented in chapter 3 which showed that further insight into the mechanisms of a vowel shift may be gained by factoring in further phonetic dimensions into the analysis. More specifically, it was shown that a case can be made for a trade-off relation between formant frequency and duration.

It seems clear that in the early sample, there is overlap in all dimensions (cf. the plots in figure 3.17 in chapter 3). However, whereas /ɛ/ overlaps almost entirely with its adjacent vowels (especially /ɪ/) in F1/F2 space, this is not the case if formant frequency is plotted against duration. In the F1/duration dimension, there is a group of /ɛ/ tokens which is clearly separated from /ɪ/. These are tokens of comparatively high duration. By implication, then, those tokens of /ɪ/ which are most distinct from /ɛ/ are of comparatively low duration, and it would be expected on the basis of the current model that further evolution of that system would favour either short /ɪ/ tokens or long /ɛ/ tokens, or both. This seems to be the case. Recall from chapter 6 that /ɪ/ is shorter in the late sample (for all /ɪ/ tokens, the difference is significant at  $p < .05$ , Wilcoxon test  $W = 3928.5$ ). In addition, it was seen that there is a range of allophones of /ɛ/ which are appreciably longer than other instances of /ɛ/, i.e. tokens occurring in monosyllables preceding a voiced stop. Therefore, the selection mechanism would favour the following variants on the basis of the distributions indicated in figure 3.17:

<i>Dimension</i>	<i>vowel</i>	<i>favoured variants</i>
F1 – F2	/ɛ/	low F1
	/ɪ/	low F1, low F2
F1 – duration	/ɛ/	high duration
	/ɪ/	low duration, low F1
F2 – duration	/ɛ/	high duration, high F2
	/ɪ/	low duration, possibly low F2

Table 8.2 – Favoured variants of KIT and DRESS tokens in the evolution of vowel systems of the type shown by the early Intermediate speakers (cf. figure 3.17, chapter 3).

As for /ɛ/, the selection mechanism would favour allophones which are comparatively close, front and long. Whereas in the F1/F2 dimension it is not clear why /ɛ/ should undergo fronting, figure 3.17 (b) shows that the longer a /ɛ/ token, the more distinct it would be from /ɪ/. Since length and fronting are correlated at this stage (cf. the distribution of /ɛ/ tokens in 3.17 (c), the correlation coefficient here is .45 at  $p < .01$ ), the development is unsurprising. With respect to /ɪ/, the selection mechanism would favour variants of low F1 and low duration in the relevant dimensions (cf. the tokens to the left of figs 3.17 (b) and (c)). If we combine these mechanisms, the centralisation of /ɪ/ is probably best explained by the fronting of /ɛ/, which implies that a high F2 is not a good predictor of ‘/ɪ/-ness’.

This in turn leaves us with the lengthening (and interrelated fronting) of a number of allophones of /ɛ/ as both the explanans and the explanandum of the change in the top two heights of the NZE short front vowel system. The causal mechanism here seems to be multidimensional as well: As /æ/ moves upward, close /ɛ/ tokens become better indicators of /ɛ/ than open ones. At the same time, the longer a token, the more distinct it is from /ɪ/. This predicts raising and lengthening of /ɛ/, but does not yet account for fronting. If we extrapolate backwards on the basis of the /ɛ/ trajectory in fig. 3.17 (c) and assume that fronting had not taken place, we would find complete overlap in the F2 – duration dimension between /ɛ/ and /æ/. Given the pattern in 3.17 (a) whereby close tokens of /ɛ/ are well separated from the distribution of /æ/, this may not sound too problematic. However, if we look at each of the three dimensions in isolation, only a combination of high duration and frontness distinguishes tokens of /ɛ/ from potential tokens of either /ɪ/ or /æ/. That is, a low F1 token can be either /ɪ/ or /ɛ/. A long token can be either /ɛ/ or /æ/. A front token can be either /ɪ/ or /ɛ/. Only long and front tokens are unambiguous instances of /ɛ/.

This would bring about a prediction that at any synchronic state of the shift, it should be long, front and close tokens of /ɛ/ which lead the shift in /ɛ/. This seems to be the case: Figure 8.1 plots /ɛ/ tokens in the early male (8.1 (a)) and early female sample (8.1 (b)), divided up into monosyllabic tokens before voiced stops and all other tokens.



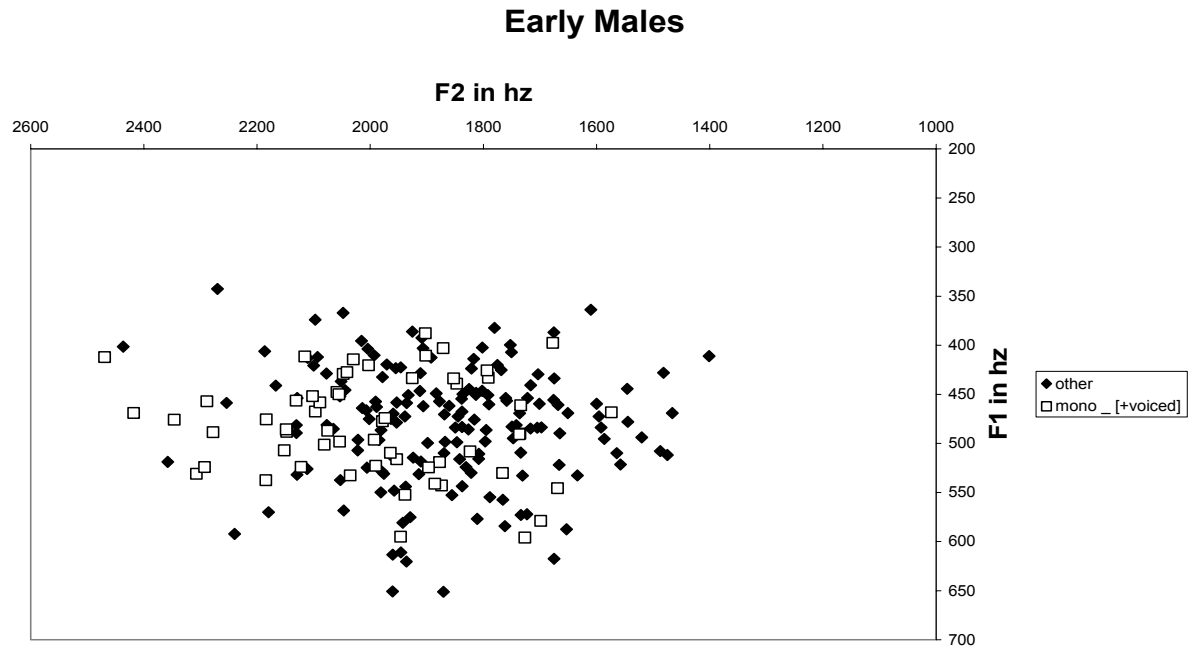


Figure 8.1 (a)

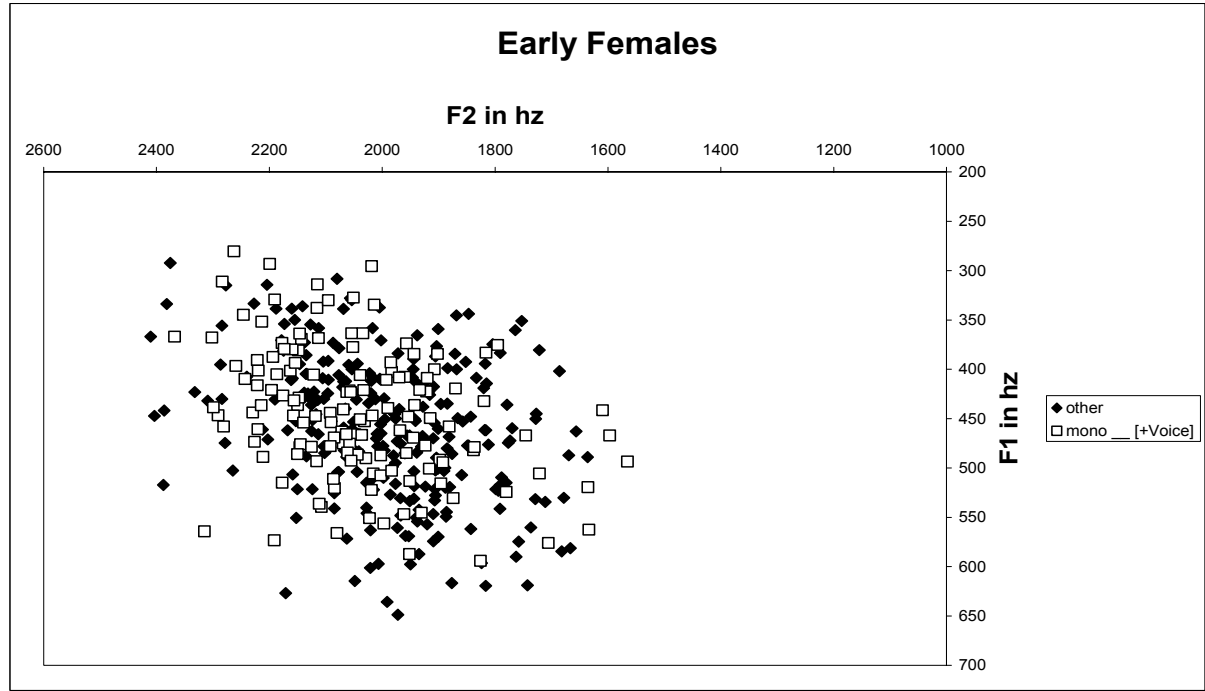


Figure 8.1 (b)

Figure 8.4 – Formant frequency plots showing tokens of DRESS both in monosyllables before voiced stops as well as their complementary environments. 8.1 (a) shows the data from the early male sample, 8.1 (b) shows early female speakers.

In the early female sample, monosyllabic tokens before voiced stops lead the shift in both spatial dimensions (Formant frequency averages of all DRESS tokens before voiced stops are 448.2 hz and 2045 hz for F1 and F2 respectively, and 460.8 hz/1994 hz before voiceless stops. A Hotelling-Lawley trace score indicates that the two distributions are significantly different below  $p < .01$ , H-L trace score = .03). In the male sample, however, there is no correlation between fronting and raising, i.e. monosyllabic /ɛ/ tokens are longer and fronter, but not significantly closer than their complementary allophones. Recall, however, that ‘long and front’ is a sufficient condition for ‘/ɛ/-ness’. Given that in the early female sample, /ɪ/ does not show fronted/raised allophones of /ɪ/ to the same extent, low F1 combined with a high F2 is a better predictor of /ɛ/.

This implies a causal loop:

1. /ɛ/ raises incipiently as a result of /æ/ raising, where the following tokens are favoured:
  - close/fronted/centralised/short/ /ɪ/. All of these increase the distance between /ɪ/ and /ɛ/.
  - fronted/long /ɛ/. Frontedness increases distance to /æ/ and some allophones of /ɪ/. Lengthening increases distance to /ɪ/.

Note that this stage, raised tokens of /ɛ/ would be more similar to /ɪ/ than non-raised tokens, all other things being equal. This does not hold for fronted/lengthened tokens: Fronting increases distance between /ɛ/ and /æ/ while maintaining distance to split /ɪ/ (i.e. an /ɪ/ distribution with both raised and centralised allophones), whereas lengthening increases distance to /ɪ/ while being approximately neutral to /æ/.

Hence, the distribution of /ɛ/ in the early male sample (fig. 8.4 (a) above) makes sense under the current model.

2. As a result, /ɪ/ raises/ fronts and centralises; /ɛ/ fronts and lengthens.
3. Since fronted and long tokens of /ɛ/ are the best predictors of /ɛ/, they become less reliable predictors of /ɪ/. As a result, /ɪ/ centralises and shortens.
4. /ɛ/ raises further. Fronting of /ɛ/ and centralisation of /ɪ/ now allow for low F1 tokens to be unambiguously categorised as /ɛ/.

This essentially amounts to claiming that the only reasonable explanation for a given language state is its immediately preceding one. The question, however, is how deterministic we want to allow such processes to be. Complete determinism seems to be ruled out on the grounds that similar inputs can bring about different results, as has happened in Australian English, for example.

Therefore, we might want to explore those stages in the process where alternative pathways of change might have been possible, and hypothetically explore these routes. The crucial point here is of course the divergent development of /ɪ/ in the two varieties. As was shown in chapter 3 above, the split system of /ɪ/ was rather predictably conditioned in that pre-velar tokens underwent raising and fronting, whereas pre-fricative tokens underwent centralisation. Since there are no velar fricatives in English, the two groups of allophones do not overlap. Any other tokens would subsequently be assigned to either one of these categories. I have proposed above that centralisation won out in NZE as a result of the selection mechanism which prefers tokens that are maximally different from the distribution of an adjacent vowel, and that /ɛ/ fronting and /ɪ/ centralisation are therefore intertwined, i.e. feed each other. Fronting of /ɛ/ itself is ‘harmless’ with regard to a split /ɪ/ vowel, as long as centralised and raised variants of /ɪ/ cancel each other out. If, however, the set of fronted/raised allophones is considerably larger than the set of centralised variants, fronting of /ɛ/ may not be neutral in this regard, provided frequency effects play a role in this type of processes. If it can be shown that AusE had, at some stage, a raised (but unfronted) /ɛ/ vowel *and* a split system where the front allophones comprise more variants than the central allophones, the model would make the right predictions. That is, whereas both pathways may

have been in principle open in both varieties, the relative timing of steps may be due to chance.

To conclude this section, the major proximate reason for why KIT centralised while DRESS fronted/raised in NZE is that, as DRESS entered the accretion zone of KIT, those DRESS tokens that were already front were those that were the least ambiguous instances of DRESS vis-à-vis both KIT (by being long) and TRAP (by being front). KIT, however, then interacts more strongly with DRESS than with TRAP, which favours the central/short variants of KIT. The prediction for a similar stage in AusE would be that one would find fewer central allophones of KIT at the crucial stage as well as a less pronounced duration difference between the central and front allophones of both KIT and DRESS.

#### 8.4 Are vowel spaces segmented?

The foregoing discussion regarding the multidimensionality of vowel spaces and the interaction between various dimensions also bears upon Labov's view regarding 'tracks' and 'adjacency on tracks' (i.e. interaction in vowel shifts is between subsets only, at least in the type of sound change discussed in this chapter). On the view proposed here, both the existence of tracks as well as the lack of interaction of elements across tracks in vowel shift and mergers may well be regarded as epiphenomenal.

Recall that the concept of 'tracks' is required in order to account for why, in vowel systems of a certain type, spatially adjacent segments do not necessarily interact with each other. However, the question is whether these segments are really adjacent if we factor in further dimensions such as duration and diphthongisation. Assume a hypothetical front vowel system with the following members:

vowel	average F1 in hz	average F2 in hz	average duration in ms
i	300	2400	200
e	400	1900	220
a	700	1500	240
ɪ	350	2000	100
ɛ	500	1700	110
ə	650	1500	120

Table 8.3 – Hypothetical front vowels and F1/F2/duration values.

We can factor in all three dimensions into a measure of distance between the vowels<sup>59</sup> as follows: Within each of the three dimensions (F1/F2/duration) involved, the respective values for each vowels are scaled within the minima and maxima within each dimension ranging from 0 to 1. For example, the F1 space ranges from 300 (0) to 650 (1), and the duration dimension ranges from 100 (0) to 240 (1). Intermediate values are fractions of 1. Distances between vowels are calculated as the Euclidean distance between two respective vowels along the three dimensions F1/F2/duration. Using this measure, the following relations hold:

<i>vowel</i>	<i>distance to</i>	<i>distance value</i>
<b>i</b>	<b>e</b>	0.63
	<b>ɪ</b>	0.85
	<b>ɛ</b>	1.13
	<b>a</b>	1.44
	<b>ə</b>	1.45
<b>ɪ</b>	<b>e</b>	0.87
	<b>ɛ</b>	0.51
	<b>a</b>	1.44
	<b>ə</b>	0.94
<b>ɛ</b>	<b>e</b>	0.85
	<b>a</b>	1.08
	<b>ə</b>	0.44

Table 8.4 – Euclidean distances between six hypothetical vowels.

<sup>59</sup> It has to be pointed out that this is a rather crude way of applying a ‘measure’ to hypothetical data. In this particular case, only a front vowel space is sampled, which implies that it involves only a fraction of F2 values one would expect in a full vowel space, which in turn leads to an overestimation of F2 in the following discussion.

Of course this model is grossly oversimplified, however, it becomes clear that if we allow for further dimensions to enter the distance equation, a number of vowels that are adjacent in F1/F2 space are rather well separated in F1/F2/duration space, from which we can infer a much simpler principle of interaction:

The interaction principle

*Adjacent vowels interact in vowel shifts*

where adjacency is defined as follows:

The adjacency principle

*Adjacency is a function of the distance between vowels along all perceivable phonetic dimensions*

These dimensions include at least the following: F1, F2, duration, and diphthongisation (i.e. amount of spectral change in F1 and F2), which would render a vowel space into 5 dimensions<sup>60</sup>. With this machinery, we can reformulate Labov's principles of vowel shifts as follows:

The vowel shift principle

*In chain shifts, vowels move in such a way as to attain a state of equidistance to adjacent vowels*

Let us now examine how this model would predict trajectories of movement in a hypothetical front vowel space. Assume an original state consisting of three long monophthongs and three short monophthongs. This situation can be diagrammed as shown in figure 8.2.

---

<sup>60</sup> Assuming that diphthongisation actually encapsulates 2 dimensions, namely spectral change in F1 and spectral change in F2, which I maintain are independent since languages are known to exhibit diphthongs that are stable in one of them (cf. Old English had 'height-neutral' diphthongs /eo/ and /æɑ/).

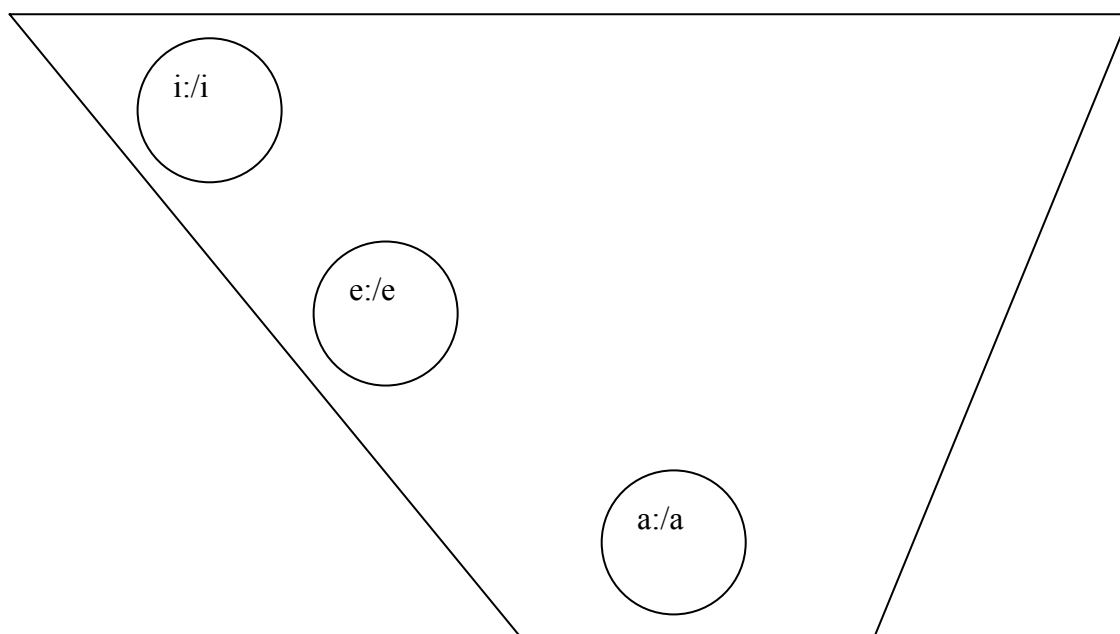


Figure 8.2 – A front-vowel space consisting of three long and three short vowels.

Assume further that within the distribution of each vowel there is a positive correlation between peripherality and duration.

Under the principle of maximising distance, we would expect the short vowels /i e a/ to move away from their respective long counterparts via the selection of variants that are maximally different from the long vowels. Since the original vowel distributions are located close to the periphery of the overall vowel space, and given a correlation between duration and peripherality, selecting short vowel variants that are maximally dissimilar to their long counterparts automatically yields movement of the short vowels away from the peripheral regions of the vowel space. This results in a system where quantity contrasts are augmented by quality differences between the originally long and short vowels (Figure 8.3):

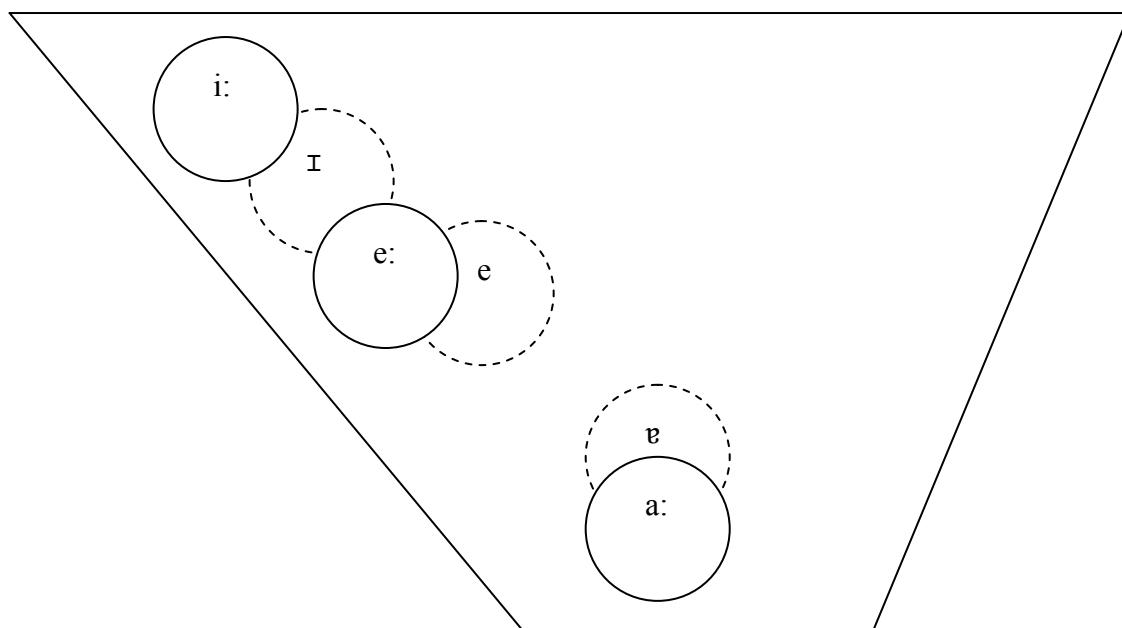


Figure 8.3 – A front vowel space with three peripheral long vowels and three non-peripheral short vowels.

Bearing in mind the claim that distances in vowel space are a function of similarity along multiple phonetic dimensions, there is likely to be a crossover point in adjacency relations where the short vowels become more similar to other short vowels than they are to their long counterparts. Therefore, chain shift relationships would be expected to be more likely to hold within each of the subsystems rather than across. That is, chain-shifting may be more likely within the sets of short or long vowels not because they are different ‘systems’ per se, but because there is greater similarity within the sets than across them. Thus, an account based purely on maximising distance might predict the emergence of ‘tracks’ along which vowels appear to move. Can it also account for the observed generalizations about how vowels tend to move along these tracks? For example, why is it that ‘lax vowels tend to fall along a non-peripheral track’ while ‘tense vowels rise along a peripheral track’? The emerged combined quantity+quality system as sketched above allows for a number of ‘permissible’ movements of different types. Consider the /e: e/ pair, for example. As the short vowel moves away from /e:/ via the selection of shorter/less peripheral variants, longer and more peripheral variants would become increasingly better predictors of /e:/. If /e:/ is mid- or closer than that,



this amounts to raising of /e:/ (by selecting those variants that are located toward the top left region of the distribution). Thus, at least for non-low vowels, this account predicts that we would generally tend to see lowering of short vowels, and raising of long vowels, as Labov predicts. As /e:/ raises, at some critical stage, /e:/ would be closer to /i:/ than to /e/, which would in turn pose an incentive for /i:/ to select variants that are better instances of /i:/. Given the state of affairs as shown above, the fittest variants would be diphthongised tokens of /i:/, since neither /e:/ nor /i/ are diphthongs. Further permissible trajectories of movement are diagrammed in figure 8.4 below, where each of the permissible pathways is indicated by an arrow. In addition, attested vowel changes following these pathways are given in brackets:

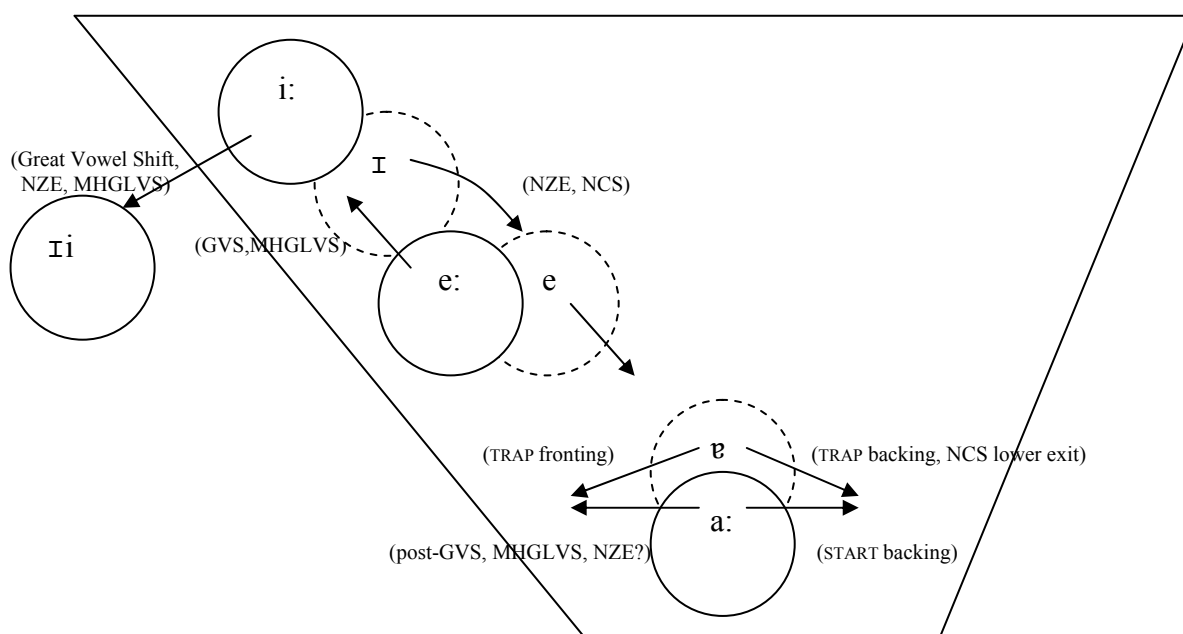


Figure 8.4 – Permissible pathways of change in a front vowel space.

The behaviour of the non-low front vowels is rather constrained. However, this is less so in the case of the central low vowels. However, once one low vowel moves into a certain direction, this would bring about constraints regarding the permissible pathways of change in the other vowel. For example, if /a:/ moves backwards, /æ/ would be expected to move forward. Once /æ/ has fronted beyond some critical threshold, /a:/ would be expected to back again. This would account for why we find the frontest realisation of START in a variety with the most raised and fronted TRAP vowel (NZE),

while back realisations of the long vowel are found in innovative RP, which also has a relatively low /ɐ/ vowel<sup>61</sup>. Again, this exemplifies the crossover in adjacency relations between vowels once some critical distance is reached.

This account predicts the types of movements that we commonly see in vowel systems. However, it is important to note that these specific movements are not incontrovertible maxims. In all cases, the exact movements the account would predict rely crucially on the initial configuration of the specific vowel space in question. Apart from being able to account for general pathways of vowel change, this approach handles a number of phenomena that remain mysterious under a model that requires explicit tracks and subsystems. I will discuss two examples in detail.

- (1) The interrelation of diphthongisation of /ei/ ('FACE') and the behaviour of the short front vowels.

I am not aware of any counterexamples to the following generalisations: (a) In varieties of English that have upward chain shifts in DRESS and TRAP, FACE is a diphthong (e.g. AusE, NZE), and (b) varieties of English that show monophthongal FACE show either downward movement of DRESS and TRAP, or downward movement of DRESS as well as upward movement *and diphthongisation* of TRAP (the Northern Cities shift, where TRAP has been shown to develop inglides as it moves upward. Labov (2001) transcribes the most innovative allophones as [ɪ:ə]). Now let us assume the above model of adjacency calculations, and add to the hypothetical front vowel space two diphthongs /ei/ and /ai/. With the addition of these elements, we need to obtain some measure of diphthongisation, which I here define as the degree of formant frequency displacement from the nucleus to the offglide from 0 (a pure monophthong) to 1 (a diphthong with the most open element /a/ as its nucleus, and the closest element as its offglide, i.e. an element of the type /ai/). Intermediate types such as /ɪi/, /ei/, /ɛi/, receive intermediate diphthongisation values between 0 and 1 that are equivalent to the proportion of the nucleus on the open-close axis from /i/ to /a/. Note that this measure is always relative to the offglide of a given diphthong. If the offglide is /a/ rather than /i/,

---

<sup>61</sup> The notational terminology is awkward since the diagram overlays an abstract representation with actual processes. The symbol /ɐ/ relates to systemic category 'short (lax) central low vowel', which in the case of English corresponds to the contemporary TRAP vowel. This vowel originated as a central vowel though.

diphthongisation values increase with closeness of the nucleus rather than openness, and a diphthong of the type /ia/ would incur a value of 1. If the offglide is not located at the extremes of the vowel space in question (an example would be the centring diphthongs in English), the diphthongisation value is intermediate. In this view, monophthongs can be conceived of as diphthongs with identical nucleus/offglide positions (e.g. /aa/, /ii/, ee/), which incur a diphthongisation value of 0.

Note that in the following illustration we are only concerned with diphthongs in the open-close trajectory. A more universally applicable formulation of this model would have to deal with both open-close diphthongs as well as front-back ones, and a combination thereof. This implies that it might be better to treat diphthongisation as two independent dimensions, namely change in the open-close dimension and change in the front-back dimension for any given vowel. However, since this is only a cursory discussion, I have decided to treat diphthongisation as a single dimension that is calculated on nucleus/offglide change in the open-close dimension only<sup>62</sup>.

For the sake of convenience, I equate the duration values of both diphthongs to that of their monophthongal counterparts.

---

<sup>62</sup> One implication of this simplified approach is that the role of diphthongisation in calculating the distance values in the following discussion is somewhat understated, since diphthongs of the type /ai/ or /ei/ show change in both the F1 as well as the F2 dimension, i.e. diphthongise in both backness and openness. Therefore, elements of the type /ai/ would receive a diphthongisation value higher than 0 in both the front-back and the open-close dimension, which would increase the distance to the pure monophthongs, which would receive a diphthongisation value of 0 in both dimensions.

vowel	to	value
<b>i</b>	<b>e</b>	0.63
	<b>ɪ</b>	0.85
	<b>ɛ</b>	1.13
	<b>a</b>	1.44
	<b>æ</b>	1.45
	<b>ei</b>	0.88
	<b>ai</b>	2.02
<b>ɪ</b>	<b>e</b>	0.87
	<b>ɛ</b>	0.51
	<b>a</b>	1.44
	<b>æ</b>	0.94
	<b>ei</b>	1.07
	<b>ai</b>	2.02
<b>ɛ</b>	<b>e</b>	0.85
	<b>a</b>	1.08
	<b>æ</b>	0.44
	<b>ei</b>	1.05
	<b>ai</b>	1.78

Table 8.5 - Euclidean distances between the six hypothetical vowels as well as two upgliding diphthongs along the dimensions F1/F2/duration/diphthongisation.

It seems clear that diphthongised /e/ (i.e. /ei/) shows a larger distance to  $\epsilon$  than the monophthong does. Further opening of the nucleus (toward /ai/), which is the case in the Australasian varieties of English, further increases the distance, and therewith the margin of security, of / $\epsilon$ / to /e/. Therefore, variants of both /e/ and / $\epsilon$ / that increase this distance would be selected, which would account for raising of DRESS as well as lowering of the nucleus of FACE in these varieties<sup>63</sup>. On the other hand, a variety that has monophthongal FACE, raising of DRESS would have the adverse effect. A similar case can be made for TRAP: If we assume a low mid-vowel for this element, and furthermore assume variation along all relevant dimensions, it seems clear that diphthongised variants are ‘fitter’ than non-diphthongised ones in a variety that has monophthongal FACE. In addition, under this account the

<sup>63</sup> This is rather simplified. Note that if the initial stage is /ei/ (for a FACE-like element) versus / $\epsilon$ / (for DRESS), lowering of the nucleus of the former vowel would actually decrease the distance between the two in the open-close dimension, while increasing the distance in the diphthongisation dimension. Therefore, lowering of the /ei/ nucleus may well be neutral with respect to the distance between the two elements, and increase that distance only at the point of equivalence of nucleus position. One implication of this line of reasoning is that DRESS raising always postdates lowering of the nucleus of FACE in varieties that have both.

diphthongisation would ‘allow’ TRAP to rise without threatening its margins of security with FACE. In a variety of the Australasian type, diphthongised TRAP tokens would be more prone to be deselected due to the presence of front centring diphthongs (which are not present in most of American English).

One important corollary of this reasoning is that interactions in a vowel space are rather complex, i.e. a vowel is sensitive to not just one of its neighbours, but to all of them (similar to the multi-body problem in celestial mechanics). The strength of the interaction of some vowel with another one is itself a function of its distance to that other vowel. This in turn handles phenomena such as

(2) The interaction of DRESS with both KIT and FLEECE in New Zealand English.

That the raising of DRESS and the centralisation of KIT are functionally related has been argued for in chapter 3 of this thesis. However, it has been claimed that DRESS also influences the behaviour of FLEECE in contemporary NZE (Maclagan and Hay (2005)). Under the principles of chain shifting as advocated by Labov, this seems rather mysterious a phenomenon given that FLEECE and DRESS belong to different subsystems. However, the account given here does not recognise subsystems in the first place (at least not in those dimensions relevant to this type of phenomenon, the phonotactics are irrelevant in this regard). Rather, the raising and lengthening of DRESS decrease the distance of that vowel to FLEECE. However, since diphthongisation is not a feature of DRESS vowel in innovative New Zealand English, diphthongised variants of FLEECE are fitter than monophthongal ones, *ceteris paribus*<sup>64</sup>.

Table 6.6 compares distance values between the three short front vowels for both subsamples broken down into two conditions, vowels in monosyllables before voiced stops and their complementary allophones. The scaling is based on pooled data (with normalised formant frequency values, and absolute duration values).

---

<sup>64</sup> It has been claimed that there was transient diphthongisation of DRESS in earlier New Zealand English. It would be interesting to see whether that diphthongisation indeed correlated negatively with FLEECE diphthongisation in some sample that includes speakers with both diphthongised and monophthongal DRESS. The account presented here of course predicts that it would.

<b>vowel</b>	<b>distance to</b>	<b>Early Males distance value</b>	<b>Late Females distance value</b>
Monosyllabic /ɪ/ /__[+voice]	Monosyllabic /ɛ/ /__[+voice]	0.29	0.45
	All other /ɛ/	0.3	0.28
	Monosyllabic /æ/ /__[+voice]	0.57	0.42
	All other /æ/	0.6	0.36
All other /ɪ/	Monosyllabic /ɛ/ /__[+voice]	0.8	0.83
	All other /ɛ/	0.16	0.38
	Monosyllabic /æ/ /__[+voice]	0.46	0.48
	All other /æ/	0.47	0.37
Monosyllabic /æ/ /__[+voice]	Monosyllabic /ɛ/ /__[+voice]	0.39	0.35
	All other /ɛ/	0.3	0.35
All other /æ/	Monosyllabic /ɛ/ /__[+voice]	0.47	0.48
	All other /ɛ/	0.33	0.4

Table 8.6 – Euclidean distances between KIT, DRESS and TRAP in the early male sample along the dimensions F1/F2/duration using normalised scales across samples. For each vowel, tokens occurring in monosyllables before voiced stops are considered separately from their complementary allophones.

The left-hand column of table 8.6 splits up each vowel into two categories: Tokens occurring in monosyllables preceding a voiced stop, and others. Column 3 shows the distance of each subsample to each other subsample. The following relations hold in the early male sample: Large distance measures hold between monosyllabic /ɛ/ before voiced stops and their *complementary* /ɪ/ tokens. This distance significantly exceeds that between other tokens of /ɪ/ and /ɛ/. On the account proposed here, we would therefore expect this part of the vowel

system to evolve in such a way that (a) those tokens of KIT and DRESS that show the greatest distance to each other are selected as the target loci of subsequent movement in each respective vowel, and (b) movement away from each other's distribution should occur to a larger extent for those tokens that show a smaller distance at the initial stage. This seems to be borne out, as the distance measures derived from the data in the late female sample (Column 4) show: The distance between the 'other' tokens of KIT and DRESS has increased more than has that of their complementary allophones.

In addition, the relation of DRESS to TRAP is not dissimilar to that between DRESS and KIT, in that those allophones which show the smallest distance to each other in the early male sample are those that move away from each other. Unfortunately, the available data points are too sparse to allow for a significant correlation between initial distance and rate of change, although a trend seems to be present. A correlation test yields a tendency ( $\rho = -.6$ ,  $p = .12$ ).

In the present formulation, the model of course makes what is in all likelihood an untenable assumption, namely that the mismatch in minimum and maximum values that we observe in comparing the two subsamples are the product of a sampling error. This is unlikely given that I showed in this thesis that in the vowel space outlined by the elements discussed here (i.e. KIT, DRESS and TRAP) change has occurred over the intermediate period in all dimensions. A proper application of the model would have to take into account an entire vowel space in order to properly scale across subsamples, which is not possible on the basis of my data (Although formant frequency measurements were taken for all vowels and speakers, duration measurements have been obtained for only a restricted subset of both, and it is likely that the short front vowels do not represent both ends of the overall duration scale).

## 8.5 Conclusions, caveats and outlook

In this chapter I proposed a model which claims that chain shifts come about as a Darwinian type of selection process that primarily aims at achieving local distance equilibria in that vowel distributions interact most strongly with distributions they are close to, in that they attempt to increase the distance (and decrease overlap, or potential confusion) between each other. This idea is similar in spirit to the model advocated by Liljencrants and Lindblom

(1972), but differs in a number of points. Firstly, all relevant phonetic dimensions that are known to play a role in vowel change need to be taken into account, which I take to be F1/F2/duration/diphthongisation, the latter encapsulating two dimensions (front-back and open-close diphthongisation). In addition, Lindblom and Liliencrants' model focuses on the outcome of evolutionary processes in vowel systems, whereas an important claim made here is that a distance model may also replicate attested pathways of change in vowel systems. Under the view proposed here, chain shifts proceed via the augmentation of differences between adjacent vowels along any one or more phonetic dimensions. The only parameter that is relevant in vowel change is distance between vowels, which is itself a function of the difference between any two elements along all of these dimensions. We do not need to invoke 'tracks' in order to account for observed patterns of interaction and can thereby avoid difficulties such as that encountered in Labov's account with regard to the short front vowel system of the Southern Hemisphere.

What this account ultimately envisages is a conception of vowel spaces as correlated fitness landscapes in the sense of Kauffman (1993, 1995), where vowel distributions are conceived of as correlates of biological organisms, which share most of their phonetic properties but vary in detail. These minor differences place them in slightly different regions of the overall landscape. Their fitness is then equivalent to their distance to members of neighbouring distributions. The fitter some variant, the higher the probability of its being selected in the subsequent evolution of the vowel system. Therefore, any given location in the fitness landscape does not by itself assign a degree of fitness to the organisms it is populated by (which would be an uncorrelated fitness landscape); rather, the fitness value thus assigned is dependent on the location of organisms of other species (the correlate to a species would here be the vowel phoneme. Any real world manifestation thereof would constitute an individual organism). Changes in the shape of the fitness landscape can occur via allophonic splits, or the borrowing of a new vowel from another language or dialect. The strength of selective pressure onto some vowel distribution is a probabilistic function of its adjacency to some other vowel distribution, i.e. vowel distributions that are closer to some distribution X than to some other distribution Y are more likely to respond to X than to Y. This also extends to variants within each of the distributions.



A model such as this is eminently testable; this, however, is far beyond the reach of this thesis. What needs to be shown is that vowel systems that have some resemblance to attested systems (i.e. simulated systems the shape and features of which are found in actual vowel systems) do indeed evolve in ways similar to their real world counterparts if modelled as proposed here. This would require detailed data from a large number of different languages that show differences in size and shape of their vowel inventory. In addition, relations between various phonetic dimensions need to be better understood. In the tentative outline implementation of the model above, I have equated the formant frequency dimension to the duration dimension (as well as the diphthongisation dimension) in a rather simplistic fashion by treating them both as linear scales, which is demonstrably wrong in the formant frequency dimension (and therewith also in the diphthongisation dimension), and has to my knowledge not been investigated in the duration dimension.

I have also deliberately shunned the discussion of a number of crucial notions such as ‘variant’ of some vowel distribution. Are these variants phonetic allophones, social allophones, or is a variant simply any utterance of some vowel? I sympathise with the last of these, however, this opens up more questions: How do we deal with misclassifications? I.e. is a token of vowel X indeed a token of X if it is perceived as Y? As a first approximation to modelling, one might simply record a number of utterances of a range of vowels in some languages and code variation along those independent variables that are the best predictors of variation. Given the data presented in this thesis (as well as in much of sociophonetics), I have reasons to assume that these factors would be phonetic ones. This might also be more prudent given that phonetic variation that is due to physiological mechanisms can be assumed to inherently skew a distribution, whereas this is not the case in social or random variation (i.e. there are coarticulatory and perceptual reasons for why vowels in rhotic environments are more central than their complementary allophones. Social variation would not produce such inherent skew).

A more fundamental problem is that such a model would presumably be incapable of modelling mergers between vowel distributions, since merging would be axiomatically ruled out. It would therefore be incapable of modelling complete attested histories of vowel change. I do not know whether a model that takes notions such as functional load into account can ever be a realistic option, as the number of factors that have been hypothesised to

play a role in the concept of ‘functional load’ is very large (Wang 1967, Labov 1994). It is also not clear whether functional load factors should best be conceived of as properties of entire distributions, or only a number of allophones, or both. In principle, however, these factors can be factored in into the model and influence the probability of interaction between vowel distributions.

I conclude that if it can be shown that modelled vowel systems adhere to principles of other complex systems, as I have tentatively done in this chapter, our view of language change would be greatly expanded in its explanatory scope, as well as open up possibilities of fruitful interdisciplinary research.

## 9.

## A Phonological Afterthought – *Vowel Change as Optimisation: Why the NZE Front Vowel Shift is not a good Example*

### 9.1 Introduction

This chapter discusses an idea put forward by Hawkins (1976) which claims that the NZE front vowel shift is an example of phonological optimisation in vowel systems. The idea of sound change as optimising phonological systems is a recurrent topic in explanatory approaches to language change.

It goes back to at least Jespersen (1909), and has been put forward on various occasions by subsequent researchers (Jakobson (1961), Martinet (1955), King (1969)). I will briefly discuss the concept of optimality in section 9.2 below. The major topic of this chapter is to present an argument against Hawkins' proposal.

### 9.2 Optimisation in phonological systems

#### 9.2.1 What is optimal?

Although it is not always clear what reflects 'optimality' in phonological systems, the following two (interrelated) points have gained some currency in the literature (I will here restrict myself to vowel systems):

(1) optimal systems are more symmetrical than non-optimal ones.

The notion of *symmetry* in vowel systems relates to either one of two interrelated things, (a) whether the front and back vowel series in a given vowel system have the same number of members in the same position, (b) whether the minimum feature specification is the same for any member of the phoneme inventory (cf. below). Consider two vowel systems comprising four and five members respectively, as schematised below.

	A			B	
	front	back		front	back
high	i	u		i	u
mid					o
low	a	ɑ		a	ɑ

In this case system A would be more symmetrical than system B, as the back vowel series in the former is a mirror image of the front vowel series. System B on the other hand is asymmetrical in that the back vowel series consists of 3 elements rather than two.

(2) optimal systems can be described with fewer features/rules.

In a similar vein, optimality has been argued to reflect a state where all members of a vowel system can be accurately specified using fewer features than is the case in some other, less optimal system. In addition, each member of a vowel system can be described by using the same features in optimal systems (i.e. the necessary number of features is sufficient to accurately specify each vowel), rather than having rogue elements that require features that are unnecessary elsewhere in the system. The two exemplary vowel systems sketched above illustrate this point as well: All members of system A can be described by two features, one height feature and one backness feature:

Feature specifications in system A

i        [+high] [-back]

a        [-high] [-back]

u            [+high] [+back]

α            [-high] [-back]

Each member of the system is therefore accurately specified, and each conceivable combination of feature value combinations is utilised by the system. System B on the other hand could be specified as follows:

i            [+high] [-back]

a            [-high] [+low]

u            [+high] [+back]

o            [-high] [-low] [+back]

α            [+low] [+back]

It is clear that in this system, only one member requires three features in order to be accurately specified, and it is only for that one vowel that an additional feature ([+/- low]) is required. Therefore, optimality in vowel systems can be regarded as a function of both spatial symmetry as well as feature matrix economy.

### 9.2.2 Hawkins' analysis of NZE vowels.

Hawkins' (1976) analysis of the short front vowel chain shift in New Zealand English argues for optimisation as the driving force behind the restructuring of the NZE short vowel space. The argument is based on a comparison of the initial and the final stage of the shift, which can be sketched as follows:

initial			final		
front	central	back	front	central	back
ɪ		ʊ	ɛ	ɪ	ʊ
ɛ		ɒ	æ	ʌ	ɒ
æ	ʌ				

The centralisation of KIT as well as the raising of DRESS and TRAP in NZE have brought about a system which is more symmetrical than the initial system as well as being more feature-economical in that (a) only two degrees of height are relevant in the shifted system (whereas three levels were required in the initial one) and (b) all possible binary feature specifications are being utilised in the final system (whereas in the initial system, the feature [+/- low] was redundant in the back series, and both height features were redundant in the central series).

### 9.2.3 General problems with the notion of ‘optimality’ in vowel systems.

General problems with the idea of optimality in sound change abound, and have been discussed *in extensu* by e.g. Lass (1980/1997) and McMahon (1994). I will give only a brief summary of these general problems, and will subsequently focus on the problems imposed upon this approach by the data analysed in this thesis.

#### *Whose symmetry?*

This relates to the fact that symmetry is essentially an aesthetic notion and rests heavily on the method of plotting. Diagrams of vowel systems that are construed on the basis of impressionistic articulatory phonetics look rather different from those based on acoustic data, which in turn look different from plots based on acoustic data, but converted onto perceptual

scales (cf. Labov 1994). From MacLagan's (1982) acoustic plots it is by no means self-evident that KIT is a central vowel for these speakers (it is clearly still closer to DRESS and TRAP than it is to STRUT in the front-back dimension), although the speakers plotted there represent fairly advanced New Zealand English. On a weighted F2 scale, KIT would be even more removed from STRUT for the same speakers.

*What about varieties that do not show the optimised set-up?*

The most vexing problem for optimality accounts of vowel change is of course the fact that so many changes that might bring about optimality do not occur in the first place, and that some changes even counteract this alleged tendency (cf. Lass' (1997) discussion on changes in the history of English verb paradigms that counteract optimisation and levelling processes). In the case of short front vowels in contemporary English, the optimal system (i.e. B above) is found only in NZE as well as in SAE, and there only under a rather favourable interpretation of the phonetic facts<sup>65</sup>. For a large majority of speakers in the UK and Northern America, however, the three-way short front vowel set-up does not seem to be much of a problem. More dramatically, the nearest relative to NZE has taken the exact opposite path: KIT has undergone fronting and raising in AusE, while the processes affecting DRESS and TRAP are rather similar. In the face of these facts, optimality can hardly be taken as a strong motivation to change vowel systems (in fact, if we accept Hawkins' diagrams and look at the speaker numbers, we might well conclude that there is a strong tendency in English to maintain non-optimal short front vowel set-ups).

However, these general points are discussed elsewhere at length. I would now like to turn to the challenge posed to optimality accounts arising from the data on the intermediate stages of New Zealand English.

---

<sup>65</sup> cf. Lass (1987) and Wells (1982) for an outline of vowel quality in SAE.

#### 9.2.4 Problems with the notion of ‘optimality’ specific to NZE short front vowels.

The major problem with accounts such as the one by Hawkins arises from a number of assumptions that are false on both theoretical as well as empirical grounds. These can be subsumed under the following themes:

##### (1) ignoring intermediates

The first problem with Hawkins’ account is the exclusive focus on the assumed initial and final stages of the short front vowel shift. However, I have shown in chapter 3 of this thesis that this vowel shift does not come about abruptly. Rather, there are intermediate stages which are properties of individuals, i.e. linguistically ‘real’ rather than merely the artefacts of plotting groups. Speaker J.M. (cf. fig. 3.16) shows such an intermediate stage, and his short front vowel system shows the following characteristics:

- DRESS and KIT show some overlap in F1/F2 space
- this overlap is not complete
- KIT tokens in velar stop environments are outside the overlapping space by being closer and/or fronter.
- KIT tokens in fricative environments are outside the overlapping space by being more central.
- most KIT tokens within the overlapping space are in bilabial stop and/or nasal environments
- DRESS tokens in nasal environments remain outside the overlapping space (by being lower)
- most DRESS tokens in the overlapping space occur in front of voiced nasals
- DRESS and KIT show almost no overlap if duration is factored in



The fact of overlap is somewhat problematic if it is to be accounted for in terms of the traditional feature framework<sup>66</sup> since things like height features specify different heights for different vowels, i.e. vowels that are different with respect to this feature should not be of the same height, which is exactly what we find for a substantial number of KIT/DRESS tokens. This might be remedied in two ways: One might either devise complex allophonic rules in order to capture the phonetic facts, or one might introduce another feature that is related to the solid duration distinction between the two. The first route would entail at least the following (grossly simplifying the phonetic facts):

/ɪ/ → [i], i.e.:	[+high]	
	[-low]	
	[+front]	
	[+raised (?)]	/ __ {k/g/ŋ}
→ [ɪ]	[+high]	
	[-low]	
	[-front]	
	[-raised]	/ __ [+cont]
		[+str]
→ [ɪ]	[+high]	
	[-low]	
	[+front]	
	[-raised]	/ __ elsewhere

Although this might work in formal terms in that one would be able to eliminate the problems of phonemically different vowels receiving identical feature specifications, the main problem with this is that it turns the notion of ‘optimality’ on its head. That is, whereas the initial stage required one additional feature to separate the front vowels, at least no further allophonic specifications are required. From the point of view of feature economy, the NZE front vowel shift moves through a stage of lesser optimality in order to eventually arrive at the optimal stage. Again, this would not be a problem if an argument

---

<sup>66</sup> Always assuming that features are indeed ‘articulatory’, or map onto articulations in a rather straightforward way rather than being merely ‘mental entities’.

could be made that such intermediate systems arise as artefacts of plotting data from different speakers, and that single speakers have either the non-optimal initial stage or the optimal final one. This, however, is not the case, which brings up the problem of teleology in sound change: Is it plausible that a generation of speakers acquire an elaborate and somewhat awkward front vowel feature specification in order to allow for their successors to have an optimal system? In addition, it seems clear that if we do want to separate those allophones of KIT that occur inside the overlapping space from those that do not, we would need a further distinctive feature. In the above feature specification I have tentatively introduced an ad hoc feature [+/- raised] which essentially fulfils that role. Again, although this might be just about permissible in technical terms, this further clutters up the phonology in that it is only necessary in order to specify a small number of allophones of one vowel only, and therefore runs counter to the alleged trend toward optimality in the system (neither the preceding nor the following stage require such a feature. No other vowel in the intermediate stage requires such a feature).

Another way out might be to simply introduce another feature in order to disentangle overlapping distributions. I showed above (cf. chapter 3) that there exist solid durational differences between KIT and DRESS. One might therefore be tempted to simply introduce a further feature (let's say [+/- long]) in order to uniquely specify each segment. Again there are both technical as well as theoretical problems with this approach. First of all, it is unclear what the actual feature would be. If we take the phonetic facts at face value, [+/-long] would probably be a plausible choice. However, recall that we are here concerned with vowels that are in the vicinity of other vowels such as FLEECE and NEAR, which are certainly much longer than both KIT and DRESS. Distinguishing KIT from DRESS by a length feature while distinguishing FLEECE from KIT and DRESS by some other feature (such as [+/- tense]) is surely *ad hoc*, since it disregards the continuum along the length dimension between the three vowels. A solution would be to recognise three degrees of length by introducing a further feature such as [+/- short]. This is questionable on theoretical grounds: First of all it seems clear that the length distinctions are solid even in the speech of speakers with little overlap between the two vowels. If features are indeed articulatory, one would have a situation in which some phonetic dimension becomes phonologically relevant only in one generation of speakers, although the nature of that

dimension does remain after the problematic stage. This brings us back to teleology: Why would a specific generation of speakers acquire a more complex feature specification than their predecessors, if the ultimate goal (as well as driving force) of this process is optimality?

(2) problems with cause-effect relationships (the push-chain mechanics)

The fact of push chain relationships between vowels that differ in some (and only that) ‘spatial’ feature (such as height or backness) is problematic for feature based accounts of phonology once we recognise the steps that comprise the overall process as *successive* events. Again, it is not the case that a speaker has either the initial or the final stage. Rather, TRAP raising and KIT centralisation occur at different points in time (as well as in space) and it is perfectly possible for one individual to show the former step, but not the latter. If we want to live without the complex machinery of allophonic sub-specification outlined above and also do not want to introduce features that are *ad hoc* as well as ephemeral, we would have to either abandon a ‘realist’ view of feature-to-articulation mappings, or choose to ignore the data (cf. the section 8.3 below).

(3) the impossibility of encoding gradualness into discrete systems

Elaborating on (2) above, it should be clear that the major problem for accounts such as Hawkins’ one is the gradual nature of this particular process (and many related processes found in other varieties of English, as well as other languages) vis-à-vis the abrupt nature of changeover in vowel typologies implicit in such feature based accounts. Any choice as to when we choose to analyse this abrupt changeover to ‘have occurred’ seems to me an arbitrary decision.

#### (4) assigning endpoints randomly

A further problem is the assignments of endpoints of such a process. More specifically, it can be asked why one would compare the vowel systems of stages A and B above and disregard not only the intermediate stages, but also the preceding as well as the following ones. For example, it seems clear that DRESS continues to rise even after KIT has moved out of the way (Maclagan and Hay 2005). Why would a process continue if the driving force behind it has vanished (i.e. the vowel system is already ‘optimal’)?

In addition, the initial stage (i.e. stage A) is a comparatively recent development in the history of English, and one may well make the argument that the preceding stages were rather more optimal than that particular typology (i.e. the short vowel system of ME was perfectly symmetrical with three back vowels and three non-back vowels, the ensuing stage can be sketched in a similar way as having three back vowels, three front vowels and one central vowel). On the whole it seems that the argument for optimality in NZE short vowels rests mainly on random choices of stages as endpoints.

### 9.3 Discussion/Conclusion

It often seems that the choice as to whether something is ‘front’, ‘back’, ‘low’, et cetera is an aesthetic choice on the part of the phonologist rather than backed by any real world manifestations of these dimensions. In fact some phonologists (e.g. Coleman 1998) clearly commit themselves to such an abstract and purely systemic view (Coleman talks about phonological representations in general rather than features). However, this does not seem to be the intention behind the original formulation of these features, which are explicitly termed ‘articulatory’ in SPE. The problem then is how we arrive at appropriate feature designations in the first place. As I pointed out before, the data presented in Maclagan (1982) shows a KIT vowel that is quite clearly closer to the other two front vowels than to STRUT. How do we know that KIT is a central vowel if there is little evidence from actual data in the case of the speakers plotted by Maclagan? Only if we accept the argument *a priori* and state that KIT is a

central vowel because if it weren't, it would receive feature specifications that are identical to DRESS.

The major conundrum seems to be that if we regard feature specifications as 'real', i.e. as mental representations that map onto articulation in a straightforward and physically specifiable sense, we cannot regard the NZE front vowel shift as an example of vowel typologies striving towards optimisation. This is due to the fact that intermediate stages with complex patterns of allophonisation are properties of individuals, and must therefore be phonologically specified. Once we do that, we lose the general tendency toward optimisation, since these intermediate stages are in many ways less optimal than both their preceding and their following ones.

In order to salvage the feature based account of the chain shift one would have to abandon the realist view and resort to a purely abstract approach to phonological features, in which features, or at least some of them, are merely 'contentless dichotomisers' (cf. Lass 1976: chapter 2, appendix) and are not related to articulatory reality at all. This, however, seems to make matters worse in that this renders real world observations of vowel articulations entirely meaningless. Therefore, decisions as to which feature specification should be assigned to any given vowel would be entirely arbitrary. This prevents us from making any statements about what is 'optimal' and what isn't, since we cannot in principle know what 'is'.

If one does accept phonetic data as phonologically relevant, one should accept all of it. In the specific case of the short front vowel shift in New Zealand English one would be obliged to take into account not only the randomly assigned endpoints, but also the typologies that are in some meaningful sense related to these stages (i.e. preceding, intermediate and following trajectories of development as well as alternative ones such as AusE, SAE, North American English and British English). Once we do this, however, the argument of phonological optimisation as a driving force behind sound change becomes implausible.

10.

## Conclusion

The major goal of this thesis was to document change in a number of phonological variables that now set New Zealand English off from those varieties it is closest to, namely Australian English, South African English, and Southern British English. The Intermediate Period, i.e. speakers born between the late 19<sup>th</sup> century and the 1930s-40s was formerly assumed to exemplify a constitutive stage in the development of contemporary New Zealand English. The results discussed in chapters 3-6 showed that this is largely correct.

Chapter 3 provided an analysis of the short front vowels /ɪ/, /ɛ/ and /æ/ from a sample of 30 speakers from the Intermediate Period. It was shown that both raising of the non-high vowels as well as centralisation of the high vowel progressed over the Intermediate Period. Whereas the early speakers, and especially the early males, show a rather conservative setup of three front vowels, medium and late speakers within this period show a system of two front vowels and a centralised vowel /ɪ/. In addition, it was shown that the developments in /ɪ/ involved a stage of rather complex allophonisation where both fronted/raised as well as relatively centralised variants were found in the speech of individual speakers. It was shown that this allophonisation was broadly conditioned by phonological environment, where centralised /ɪ/ occurred mainly in fricative environments, while the fronted/raised variants tend to occur in velar environments. CART analysis revealed that for all three front vowels, phonemic environment tends to be a better predictor of formant frequency than speaker sex or age, i.e. internal factors override external ones.

As for the mechanisms of the overall shift a case was made for a push chain scenario, because we find speakers with raised /ɛ/ but uncentralised /ɪ/, while the reverse doesn't hold. Since this relationship is a property of individual speakers rather than an artefact of plotting pooled data, we have evidence that to some extent at least, push chains come about sequentially, i.e. it is not a requirement that any given individual should show either the conservative or the innovative pattern. This finding suggests a stage where we find (partial) merger between the distributions of /ɪ/ and /ɛ/ with a large degree of overlap between the two, and subsequent unmerging. However, it was further shown that once duration is factored in as a variable, the two vowels remain rather distinct even for speakers who show a large degree of overlap in the F1/F2 dimension. The analysis of duration revealed that /ɪ/ is much shorter than both /ɛ/ and /æ/, and that durational allophonisation is highest in /ɛ/. From this it was hypothesised that duration may well be a potent cue in distinguishing the short front vowels for these speakers. This, in turn, has implication for our conceptualisation of the relevant dimensions in the structuring of the front vowel space in English, a topic I took up again in chapters 7-9.

A phonemic merger rather than a chain shift was the topic of chapter 4. Whereas contemporary speakers of NZE tend to show merger between /ɪə/ and /ɛə/, the earliest speakers analysed by Gordon et al. (2004) show no such merger. It was therefore a reasonable hypothesis to date the onset of the merger back to the Intermediate Period. It was found that whereas for the most part, Intermediate speakers keep the two phonemes apart, there is a clear tendency of a development toward eventual merger, during which the two vowels approximate each other over the Intermediate Period. A somewhat weaker case was made for spatial expansion of both vowel phonemes, which may suggest a combination of both merger-by-approximation as well as merger-by-expansion in the overall process of the /ɪə/-/ɛə/ merger in NZE. An interesting detail that emerged from measuring the distances between /ɪə/ and /ɛə/ for the six subsamples was that if one resolves the data for different degrees of stress, one finds that the most innovative speakers (the late females) show the smallest distance (and therewith the highest similarity) between /ɪə/ and /ɛə/ in environments of high stress, which might be indicative of a change-over in the cognitive representation of the two vowels (which would now be one phonological element).

Furthermore, chapter 5 raises the issue of the relation between the diphthong and the monophthong system. This is an important point and has both theoretical as well as practical implications<sup>67</sup>, and it was shown that the relations are far from clear for these speakers. A number of different measures were employed to assess these relations. A comparison of formant frequency point measurements showed that some speakers show congruence between the nucleus of /ɛθ/ and the /ɛ/ vowel in F1/F2 space, while others keep them distinct. A dynamic measure indicates that if we overlay the duration of the monophthongs /ɪ/ and /ɛ/ onto the trajectories of the front centring diphthong, we arrive at vastly different diphthong shapes for the two diphthongs. That is, whereas /ɪ/ is considerably shorter than /ɛ/, /ɪθ/ and /ɛθ/ do not show such a duration difference. A similar conclusion was reached by comparing not the overall duration, but the progression-through-overall-frequency-range of the diphthongs on the basis of the monophthongal duration values. From this I concluded that whereas a case may be made for equating the nucleus of /ɛθ/ with /ɛ/, we can rule out congruence between the nucleus of /ɪθ/ and /ɪ/ (and would therefore have to conclude that a transcription such as /ɪθ/ is inadequate in the case of NZE speakers from at least this stage onward). If the nucleus of /ɪθ/ has an equivalent in the monophthong system, it would be /i/ (FLEECE). Again, however, this is not a clear cut case since some speakers keep the two distinct, and one might have to accept the conclusion that, in terms of their phonetic properties at least, diphthongal nuclei may not have direct equivalents in the monophthong system.

The third analytical chapter (chapter 6) looked at the development of so-called ‘broad A’ in the intermediate Period. While contemporary NZE is a full broad A variety (i.e. any word that contains broad A in any broad A variety contains broad A in NZE), Gordon et al. showed that this was not the case for MU speakers. Rather, these speakers often retain flat-A, and this retention is more likely to occur in pre-nasal rather than pre-fricative environments. The data analysed in chapter 6 corroborates this finding further in that early Intermediate speakers show more flat A than late speakers do, and that flat A is more likely to occur in pre-nasal environments. This constitutes further evidence that NZE inherited the broad A change in progress rather than the completed system. Distributional data additionally showed that the process comes about by lexical transfer rather than gradual shift. Finally, there is

---

<sup>67</sup> Practical consequences would include questions of how to appropriately transcribe diphthongal nuclei, and whether diphthongal nuclei should be included in studies of their (assumed) monophthongal equivalences.



cursory evidence that proper nouns such as place names or family names are more likely to retain the conservative variant. The data, however, is rather too scarce to make a clear-cut case for this.

In addition to presenting an analysis of these variables I discussed a number of theoretical implications for the study of vowel change. In chapter 7, I discussed the alleged exceptional status of the short front vowels in Labov's theory of directionality in vowel shifts. The main line of reasoning was that because the English front vowel space is rather untypical of complex vowel systems, its members may not be good examples of certain types of vowels, and would hence not be expected to show typical properties of these vowels. More specifically, I argued that the merger of /ɛ:/ and /e:/ on /i:/ that occurred in EModE as well as the fronting of ME /a/ created a front vowel space that shows no long mid-vowels, but two short ones (at least from an etymological point of view). This would therefore allow language learners to reinterpret the formerly short (or 'lax') vowels as long (or 'tense') ones. These elements are therefore indeterminate due to the lack of reference points in their immediate vicinity (whereas such reference points exist for the other 'lax' vowels). Therefore, the lack of directionality in the development of these vowels across different varieties of English is not unexpected at all. This hypothesis is substantiated by two facts, namely (1) the lowering of /ɛ/ in varieties with an etymological 'gap-filler', usually monophthongal /eɪ/, which would then allow the closest short vowel to be unambiguously classified as a short vowel, and (2) phonetic data from the Intermediate Period analysed in chapter 3, which suggests that the durational properties of /ɛ/ and /æ/ are rather more similar to that of long vowels than those of /ɪ/ are in terms of both average duration as well as durational allophonisation (the latter being a property of /ɛ/ especially). The bottom line was to regard tracks and peripherality relations as concepts that may not necessarily hold across entire vowel spaces in a timeless fashion with elements switching their allegiances to given tracks over time, but rather are reassessed during the course of sound change by individual speakers.

Chapter 8 went one step further and claimed that the concept of 'tracks' may better be abandoned altogether. I set out by proposing a rather simple evolutionary mechanism that may help explain directionality in chain shifts over time and that would furthermore be able to explain this directionality. The model is functional but not fully deterministic, i.e. it is historical. The essential claim was that chain shifts proceed via the selection of fitter variants

from an overall pool of utterances of the vowel phonemes involved. Fitness is ultimately a function of distance to all other vowel distributions. Distance is in turn a function of similarity to other vowel utterances along all relevant phonetic dimensions, i.e. a given vowel space is multidimensional at all times, and distance is calculated on the basis of at least five dimensions (F1/F2/duration/diphthongisation-on-the-vertical (i.e. F1)/diphthongisation-on-the-horizontal (F2)). Once we incorporate dimensions other than the purely spatial ones of backness and openness, we can conceive of chain shifts as functional relationships between truly adjacent vowels without having to admit apparent violations such as the mid-vowels having no effect on the high short vowels in the GVS, although these are certainly more adjacent on a view that recognises only the spatial dimensions. On this view, the high short vowels are simply much further away from the long mid-vowels in multidimensional phonetic space than the high long vowels are. On this view, further violations of Labov's theory can be explained away. An example I discussed is the interaction between /ɛ/ and /i:/ in contemporary New Zealand English, which would be unexpected on the basis of the Labovian view of vowel systems (since the two belong to different 'subsystems'). The model as proposed here does not recognise subsystems that impose a priori structure onto a given vowel space (phonotactics is a different matter altogether and in my opinion irrelevant to this discussion). Rather, subsystems arise as posthoc generalisations by the analyst on a binary axis of phonetic classification.

What has to be pointed out in this context is that the model is not deterministic in any strict sense of the word. Rather, at any given point in time, further evolution is only vaguely directional. I exemplified this by comparing the differential effects of the raising of /ɛ/ onto the subsequent evolution of /ɪ/ in AusE and NZE respectively. Given initial raising of /ɛ/ in both varieties (or in the common ancestor of the two), both raised and centralised variants of /ɪ/ would receive a higher fitness value than the original realisation, and that is all that matters. Whether centralisation, fronting, or both (as in SAE) win out, is a matter of contingency. However, once a particular path is chosen, a reversal is irreversible under this view (i.e. fronting of /ɪ/ in present day NZE would be entirely unexpected, all other things

being equal<sup>68</sup>). Finally, the strength of the interaction of a given vowel with another vowel is a *gradual* function of distance to that vowel.

I adduced some data from the Intermediate speakers to test a number of predictions of the model, and although this was done in a highly idealised fashion, a number of predictions were borne out (e.g. allophones of a given vowel phoneme of comparably low fitness move toward the vowel space occupied by allophones of higher fitness, and change *faster* than those of high fitness). I pointed out that in order to properly test such a model one would have to amass much larger amounts of data from a wide selection of vowel systems of different complexity. I also brought up a number of caveats, the most important of which is the incapability of this view to account for merger between two vowels. This may, however, be overcome by incorporating auxiliary concepts such as ‘informative load’ into a fitness calculation.

In chapter 9 I discussed an earlier idea that the NZE front vowel shift is an example of phonological optimisation in vowel systems. I argued this idea is untenable if one takes into account the data from the Intermediate period. More specifically, the optimisation argument would require a certain generation of speakers to ‘sacrifice’ optimality in their short vowel phonology in order to allow for an optimal system to eventually develop. If we try to feature-specify the short front vowels of early Intermediate speakers, we arrive at exceedingly complex mappings from phoneme to allophone under the classic SPE framework. A further argument against the idea of optimisation was made on the basis of cross-dialectal comparison, where it was concluded that even if we regard the current NZE short vowel system as ‘optimal’, it certainly isn’t popular, in that most other varieties do not have it. This, however, would be expected if chain shifts are driven by a tendency to feature-economy and spatial symmetry in vowel systems.

Apart from the empirical acoustic analysis and the theoretical discussion I decided to incorporate a somewhat lengthy methodological discussion into the thesis in chapter 2. I compared three commonly used normalisation methods (Neary’s log mean method, Gerstman’s range method, and Lobanov’s z-score derivation) and attempted to evaluate them using a number of data pools from the Intermediate sample. I concluded on the somewhat

---

<sup>68</sup> ‘All other things being equal’ here means that all other vowels do not evolve any further. It is, however, possible to devise a trajectory of future development where re-fronting of /ɪ/ would be expected.

bleak note that normalisation is both necessary as well as ultimately impossible if one adheres to the conventional view of representing group plots in F1/F2 charts. It is necessary since the only alternative would be to show only individual vowel systems, which would make any large-scale studies cumbersome in both analysis and presentation. It is impossible (at least if one attempts to achieve ‘perfect’ normalisation) since it is strictly speaking circular: In order to find one or more calibration points in a given vowel space, one needs to know whether a given vowel is stable in a speech community or not. This, however, is *always* an open question and not an a priori given. Former attempts to handle this problems have resorted to either auditory cross-checking (Disner 1980), maximal control of social variables (Adank et al. 2005), or a priori assumptions about what normalised data should correlate with (Hindle 1978). None of these is entirely satisfactory for reasons discussed in chapter 2. I concluded that researchers on vowel change might have to resort to stating distance values between vowels in the future rather than rely on formant frequency plots and state vowel change as change in distance differentials between three or more vowels over time. This, however, might be an opportunity rather than a loss: I argued in chapter 8 that distances and adjacency relations in multidimensional vowel spaces go a long way in explaining functional relationships between vowels in push chains, and a representation along these lines might therefore allow for analysing vowel spaces irrespective of physiological factors and arrive at the functional relationships at the same time.

Of course, this opens up more questions than it answers with respect to the practical feasibility of such an approach, but this would not be an academic thesis if it did not.

## References

Adank, P. (2003) *Vowel normalisation: A perceptual-acoustic study of Dutch vowels*.  
Doctoral dissertation, Nijmegen.

Adank, P., van Houdt, R., and Smits, R. (2004) 'An acoustic description of the vowels  
of Northern and southern Standard Dutch.' *Journal of the Acoustical Society of  
America* 116: 1729-1738.

Aitken, A.J. (1981) 'The Scottish vowel length rule.' In M. Benskin et al. (eds.): *So many  
people, longages and tonges: Philological essays in Scots and medieval  
English presented to Angus McIntosh*. Edinburgh: The Middle English Dialect  
Project.

Baayen, R.H., Piepenbrok, R. and Gulikers, L. (1995) The CELEX lexical database (CD  
ROM). Philadelphia: University of Pennsylvania.

Bartlett, C. (1992) 'Regional Variation in New Zealand English: The Case of  
Southland.' *New Zealand English Newsletter* 6: 5-15.

- Batterham, Margaret (1995) *Some Front Vowel Variables in New Zealand English*.  
Doctoral dissertation, Melbourne: LaTrobe.
- Bauer, Laurie (1979) 'The Second Great Vowel Shift'. *Journal of the International Phonetic association* 9.2: 57-66.
- Bauer, Laurie (1986) 'Notes on New Zealand English Phonetics and Phonology'.  
*English World-Wide* 7: 225-258.
- Bauer, Laurie (1992) 'The Second Great Vowel Shift Revisited'. *English World-Wide*  
13.2: 253-268.
- Benediktsson, H. (1970) *The Nordic languages and modern linguistics*. Reykjavik:  
Societas Scientarium Islandica.
- Bernard, John R. (1970) 'Toward the acoustic specification of Australian English'.  
*Zeitschrift für Phonetik und Kommunikationswissenschaft* 23: 113-128.
- Blevins, J. (2004) *Evolutionary Phonology: The emergence of sound patterns*.  
Cambridge University Press.
- Breiman, L., Friedman, J., Olshen, R. & Stone, C. (1984) *Classification and Regression Trees*. New York: Chapman & Hall.
- Bybee, J. (2001) *Phonology and language use*. Cambridge University Press.
- Bybee, J. and Hopper, P. (eds, 2001) *Frequency and the emergence of linguistic structure*. John Benjamins.

- Chistovich, L.A. and Lublinskaya, V.V. (1979) 'The centre-of-gravity effect in vowel spectra and critical distance between the formants'. *Hearing Research* 1: 185-195.
- Chomsky, N. and Halle, M. (1968) *The sound pattern of English*. New York: Harper and Row.
- Cleveland, W. S. (1979) 'Robust locally weighted regression and smoothing scatterplots'. *Journal of the American Statistical Association* 74: 829-836.
- Coleman, J.S. (1998) *Phonological representations: Their names, forms, and powers*. Cambridge University Press.
- Cox, F. (1996) *An acoustic study of vowel variation in Australian English*. Doctoral dissertation, Sydney: Macquarie University.
- Croft, W. (2000) *Explaining language change: An evolutionary approach*. New York: Longman.
- Crystal, T.H. and House, A.S (1982) 'Segmental durations in connected speech signals: Preliminary results.' *Journal of the Acoustical Society of America* 72: 705-716.
- Crystal, T.H. and House, A.S (1988 (a)) 'Segmental durations in connected speech signals: Current results.' *Journal of the Acoustical Society of America* 83: 1553-1573.
- Crystal, T.H. and House, A.S (1988 (b)) 'The duration of American English vowels: An overview.' *Journal of Phonetics* 16: 263-284.

- de Boer, B. (2000) 'Emergence of sound systems through self-organisation.' In C. Knight et al. (eds.): *The evolutionary emergence of language: Social function and the origins of linguistic form*. Cambridge University Press.
- de Boer, B. (2001) *The origin of vowel systems*. Oxford University Press.
- Disner, S. (1980) 'Evaluation of vowel normalisation procedures.' *Journal of the Acoustical Society of America* 67: 253-261.
- Dobson, E.J. (1957) *English Pronunciation 1500-1700*, Vol. 2: *Phonology*. Oxford University Press.
- Fant, G. (1960) *Acoustic theory of Speech Production*. The Hague: Mouton.
- Fant, G. (1973) *Speech sounds and features*. MIT Press.
- Fujimura, O. (1962) 'Analysis of nasal consonants.' *Journal Of The Acoustical Society Of America*. 34: 1865-1875.
- Fujisaki, H. and Kawashima, T. (1968) 'The roles of pitch and higher formants in the perception of vowels.' *IEEE transactions on audio and electroacoustics* AU 16: 73-77.
- Gerstman, L. (1968) 'Classification of self-normalised vowels.' *IEEE transactions on audio and electroacoustics* AU 16: 78-80.
- Glasberg, B.R. and Moore, B.J.C. (1990) 'Derivation of auditory filter shapes from noise-notched data.' *Hearing research* 47: 103-138.



- Gordon, E. and MacLagan, M. (1989) 'Beer and bear, cheer and chair: A longitudinal study of the ear/air contrast in New Zealand English'. *Australian Journal of Linguistics* 21.2: 215-238.
- Gordon, E. and MacLagan, M. (2001) "'Capturing a sound change": a real time study over 15 years of the NEAR/SQUARE diphthong merger in New Zealand English'. *Australian Journal of Linguistics* 9: 203-220.
- Gordon, Elizabeth, L. Campbell, J. Hay, M. MacLagan, A. Sudbury & P. Trudgill (2004) *New Zealand English: Its origin and evolution*. Cambridge University Press.
- Gordon, Matthew J. (2001) *Small-Town values, big city vowels: A study of the Northern Cities shift in Michigan*. Durham: Duke University Press.
- Gordon, Matthew J. (2002) Investigating Chain Shift and Mergers. In J.K. Chambers et al. (eds.): *The Handbook of Language Variation and Change*. Blackwell.
- Harrington J., Palethorpe, S., & Watson, C. (2000 (a)), 'Does the Queen Still Speak the Queen's English?'. *Nature* 407: 927-928.
- Harrington J., Palethorpe, S., & Watson, C. (2000 (b)), 'Monophthongal vowel changes Received Pronunciation: an Acoustic Analysis of the Queen's Christmas Broadcasts.' *Journal of the International Phonetic Association* 30: 63-78.
- Harrington, Jonathan & Cassidy, Steve. (1999) *Techniques in speech acoustics*. Boston : Kluwer Academic Publishers.
- Hawkins, P. (1976) 'The role of NZ English in a binary feature analysis of English short vowels.' *Journal of the International Phonetic Association* 6.2: 50-66.

- Hay, J. and Bresnan, J. (2006) 'Spoken syntax: The phonetics of 'giving a hand' in New Zealand English.' *The Linguistic Review* 23: 319-347.
- Hindle, D. (1978) 'Approaches to vowel normalisation in the study of natural speech.' In Sankoff (ed.): *Linguistic variation: Models and methods*. New York: Academic Press.
- Holmes, J. and Bell, A. (1992) 'On shear markets and sharing sheep: The merger of EAR and AIR diphthongs in New Zealand English'. *Language Variation and Change* 4: 251-273.
- Jacques, R. and Rastatter, M. P. (1990). 'Recognition of speaker-age from selected acoustic features as perceived by normal young and geriatric listeners.' *Folia Phoniatrica* 48: 118-124.
- Jakobson, R. (1961) 'The concept of the sound law and the teleological criterion'. In *Selected writings*. The Hague: Mouton.
- Jespersen, O. (1909) *A modern English grammar on historical principles*. Copenhagen: Munksgaard.
- Johnson, K. (1989) 'Higher formant normalization results from auditory integration of F2 and F3'. *Perception and Psychophysics* 46:174-180.
- Johnson, K. (1990) 'Contrast and normalisation in speech perception'. *Journal of Phonetics* 18:229-254.
- Jones, W. (1786) The third anniversary discourse. In Harris (ed. (1993)): *Discourse delivered at the Asiatick Society*. London: Routledge.

- Kauffman, S. (1993) *The origins of order: Self-organisation and selection in evolution*. Oxford University Press.
- Kauffman, S. (1995) *At home in the universe: The search for laws of self-organisation and complexity*. Oxford University Press.
- Keller, R. (1994) *On language change: The invisible hand in language*. London: Routledge.
- King, R.D. (1969) *Historical linguistics and generative grammar*. Englewood Cliffs: Prentice-Hall.
- Kiparsky, P. (1995) 'The phonological basis of sound change.' In Goldsmith (ed.): *The Handbook of Phonological Theory*. Blackwell.
- Klatt, D.H. (1973) 'Interaction between two factors that influence vowel duration.' *Journal of the Acoustical Society of America* 55: 1102-1104.
- Kökeritz, H. (1953) *Shakespeare's pronunciation*. New Haven: Yale University Press.
- Labov, W. (1981) 'Resolving the neogrammarian controversy'. *Language* 57:267-309.
- Labov, W. (1994) *Principles of Linguistic Change*, Vol. 1: *Internal Factors*. Blackwell.
- Labov, W. (2001) *Principles of Linguistic Change*, Vol. 2: *Social Factors*. Blackwell.
- Labov, W., Yaeger, M., and Steiner, R. (1972) *A quantitative study of sound change in progress*. Report on National Science Foundation Contract NSF-GS-3287, University of Pennsylvania.

Langstrof, C. (2004) 'The centring diphthongs in intermediate New Zealand English – an acoustic analysis.' Proceedings of the 10<sup>th</sup> Australian International Conference on Speech Science and Technology. Published on CD-ROM: pp 207-212.

Lass (1976) *English phonology and phonological theory*. Cambridge University Press.

Lass (1980) *On explaining language change*. Cambridge University Press.

Lass (1984) *Phonology: An introduction to basic concepts*. Cambridge University Press.

Lass (1987) *The shape of English: Structure and history*. Cambridge University Press.

Lass (1990) 'Where do extraterritorial Englishes come from? Dialect input and recodification in transported Englishes'. In Adamson et al. (eds): *Papers from the fifth international conference on English historical linguistics*. Amsterdam: John Benjamins.

Lass (1992a) 'What, if anything, was the Great Vowel Shift?' In M. Rissanen et al. (eds.): *History of Englishes : New methods and interpretations in historical linguistics*. Berlin: Mouton de Gruyter.

Lass (1992b) 'Phonology and Morphology'. In N.F. Blake (ed.): *The Cambridge history of the English language*, Vol. 1: 1066-1476. Cambridge University Press.

Lass (1997) *Historical linguistics and language change*. Cambridge University Press.

Liberman, A. (1996) *Speech: A special code*. MIT Press.

Liljencrants, J. and Lindblom, B. (1972) 'Numerical simulation of vowel quality systems: The role of perceptual contrast.' *Language* 48: 839-863

- Lindblom, B. (1984) 'Can the models of evolutionary biology be applied to phonetic problems?'. In M.P.R. van den Broecke and A. Cohen (eds.): *Proceedings of the 10<sup>th</sup> international congress of phonetic sciences*. Dordrecht: Foris.
- Lindblom, B. (1998) 'Systemic constraints and adaptive change in the formation of sound structure.' In J.R. Hurford et al (eds.): *Approaches to the evolution of language*. Cambridge University Press.
- Linville, S.E. and Fisher, H.B. (1985) 'Acoustic characteristics of perceived versus actual vocal age in controlled phonation by adult females.' *Journal of the Acoustical Society of America* 78: 40-48.
- Lobanov, B. (1971) 'Classification of Russian vowels spoken by different speakers'. *Journal of the Acoustical Society of America* 49: 606-8.
- Luce, P.A. and Luce, J.C. (1985) 'Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production'. *Journal of the Acoustical Society of America* 78: 1949-1957.
- Luick (1914) *Historische Grammatik der Englischen Sprache*. Leipzig: Tauchnitz.
- Jespersen, O. (1909) *A modern English grammar on historical principles*, Vol. 1: *Sounds and spelling*. Copenhagen: Munsgaard.
- MacLagan, M. (1982) 'An acoustic study of New Zealand English vowels'. *The New Zealand Speech Therapists' Journal* 37.1: 20-26.
- MacLagan, M. and Gordon, E. (1996) 'Out of the AIR and into the EAR: another view of the New Zealand diphthong merger'. *Language Variation and Change* 8: 125-147.

MacLagan, M. and Gordon, E. (2000) 'The NEAR/SQUARE merger in New Zealand English'. *Asia Pacific Journal of Speech, Language and Hearing* 5.3: 201-207.

MacLagan, M. and Hay, J. (2005) 'Getting fed up with our feet: Contrast Maintenance and the New Zealand English 'short' front vowel shift.' To appear in *Language Variation and Change*.

Mckenzie, J. (2005) DRESS and FLEECE in modern New Zealand English. Unpublished honours extended essay, University of Canterbury.

MacMahon, A. (1994) *Understanding language change*. Cambridge University Press.

Martinet, A. (1955) *Économie des changements phonétiques : traité de phonologie diachronique*. Berne : Francke.

Mendoza-Denton, Norma, Hay, J., Jannedy, S. (2003). Probabilistic Sociolinguistics: Beyond Variable Rules. In R. Bod, J. Hay & S. Jannedy (eds.), *Probabilistic Linguistics*. Cambridge: MIT Press.

Miller, R.L. (1953) 'Auditory tests with synthetic vowels.' *Journal of the Acoustical Society of America* 25: 114-121.

Miller, J.D. (1989) 'Auditory perceptual interpretation of the vowel.' *Journal of the Acoustical Society of America* 85: 2114-2134.

Moon, S.J. and Lindblom, B. (1994) 'Interaction between duration, context, and speaking style in English stressed vowels.' *Journal of the Acoustical Society of America* 96: 40-55.

- Moore, B.C.J. and Glasberg, B.R. (1983) 'Suggested formulae for calculating auditory-filter bandwidths and excitation patterns.' *Journal of the Acoustical Society of America* 85: 365-378.
- Neary, T.M. (1978) *Phonetic feature systems for vowels*. Indiana University Linguistics Club.
- Neary, T.M. (1989) 'Static, dynamic, and relational properties in vowel perception'. *Journal of the acoustical Society of America* 85: 2088-2113.
- Nordström, P.E. and Lindblom, B. (1975) 'A normalisation procedure for vowel formant data.' *Proceedings of the 8<sup>th</sup> International Congress of Phonetic Sciences*.
- Ohala, J.J. (1989) 'Sound change is drawn from a pool of synchronic variation'. In L.E. Breivik and E.H Jahr (eds.): *Language change: contributions to the study of its causes*. Berlin: Mouton de Gruyter.
- Paul, H. (1888) *Principles of the history of language*. London: Swan Sonnenschein.
- Peterson, G.E. and Barney, H.L. (1952) 'Control methods used in the study of vowels.' *Journal of the Acoustical Society of America* 24: 175-184.
- Peterson, G.E. and Lehiste, I. (1960) 'Duration of syllable nuclei in English'. *Journal of the Acoustical Society of America* 32: 693-703.
- Pinker, S. (1995) *The language instinct*. Penguin Books.
- Popper, K. (1982) *Logik der Forschung*. Tübingen: Mohr.

- Port, R.F. (1981) 'Linguistic timing factors in combination'. *Journal of the Acoustical Society of America* 69: 262-?.
- Priebsch, R. and Collinson, W. (1958) *The German language*. London: Faber & Faber.
- Rastatter, M.P., McGuire, R. A., Kalinowski, J. and Stuart, A. (1997) 'Formant frequency characteristics of elderly speakers in contextual speech.' *Folia Phoniatrica et Logopedica* 49: 1-8.
- Stevens, K. and House, A. (1963) 'Perturbation of vowel articulations by consonantal context: An acoustical study'. *Journal of Speech and Hearing Research* 6: 111-128.
- Stevens, S.S., Newman, E.B. and Volkmann, J. (1937) 'A scale for the measurement of the psychological magnitude pitch.' *Journal of the Acoustical Society of America* 8: 185-190.
- Stockwell, R. (1972) 'Problems in the interpretation of the Great Vowel Shift'. In M.E. Smith (ed.): *Studies in linguistics in honour of George L. Trager*. The Hague: Mouton.
- Syrdal, A.K and Gopal, H.S. (1986) 'A perceptual model of vowel recognition based on the auditory representation of American English vowels'. *Journal of the Acoustical Society of America* 79:1086-1100.
- Trager, G. and Smith, H.L. (1957) *An outline of English structure*. Washington D.C.: American Council of Learned Societies.
- Trudgill, P. (1986) *Dialects in contact*. Blackwell.



- Trudgill, P. (2004) *New dialect formation: The inevitability of colonial Englishes*. Oxford University Press.
- Trudgill, Peter, E. Gordon & G. Lewis (1998) 'New Dialect Formation and Southern Hemisphere English: The New Zealand Short Front Vowels'. *Journal of Sociolinguistics* 2.1: 35-51.
- Umeda, N. (1975) 'Vowel duration in American English'. *Journal of the Acoustical Society of America* 58: 434-445.
- Verner, K. (1875) 'Eine Ausnahme der ersten Lautverschiebung'. *Zeitschrift für vergleichende Lautforschung auf dem Gebiete der indogermanischen Sprachen* 23: 97-130.
- Warren, P. (2006) 'Word recognition and sound merger.' In June Luchjenbroers (ed.): *Cognitive linguistics investigations*. John Benjamins. pp 169-186.
- Warren, P., Hay, J., and Thomas, B. (2004) 'The loci of sound change effects in recognition and perception.' Paper presented at the ninth conference on Laboratory Phonology in Champaign-Urbana in June 2004.
- Wang, S.-Y. (1967) 'The measurement of functional load'. *Phonetica* 16: 36-54.
- Wang, S.Y. (1977) *The lexicon in phonological change*. The Hague: Mouton.
- Wells, J. C. (1982) *Accents of English*. Cambridge University Press.
- Wright, J.T. (1986) 'The behaviour of nasalised vowels in the perceptual vowel space'. In: J.J. Ohala and J.J. Jaeger (eds.), *Experimental Phonology*. New York: Academic Press.

Zwicker, E. and Terhardt, E. (1980) 'Analytical expressions for critical-band rate and critical bandwidth as a function of frequency.' *Journal of the Acoustical Society of America* 68: 1623-1625.